



UNIVERSITAT_{DE} BARCELONA

Final Degree Project

Biomedical Engineering Degree

**Evaluation with an Independent Dataset
of a Deep Learning-based Left Atrium
Segmentation Method**

Barcelona, 14 of June 2021
Author: Andrea Nadal Pellisé
Director/s: Gaspar Delso
Roser Sala
Tutor: Manel Puig

ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisors Roser Sala and Gaspar Delso for their continuous support along the execution of this thesis and for giving me the opportunity to take part in such an interesting project. I am sincerely grateful for their availability and their commitment to the project.

My deep gratitude also goes to Jana Reventós, without whom this work would not have been possible. She provided me with her Master Thesis paper and code, and always showed genuine willingness to help me.

Additionally, I would like to thank my colleagues Laia, Judit, Lucía, and Maria for helping me solve some technical issues that emerged during the development of the project, and Esther, for her empathy and support in very demanding moments.

Also, this work would not have been possible without the Arrhythmias Section at Hospital Clínic de Barcelona, who provided the data needed for the research process.

Last but not least, I would like to thank my family and friends for having encouraged me throughout the whole degree, and especially throughout this Final Degree Thesis.

ABSTRACT

Atrial fibrillation (AF) is the most prevalent type of arrhythmia nowadays. Even though it is associated with significant morbidity and mortality, there is still a substantial lack of basic understanding of the left atrium (LA) and pulmonary veins (PVs) anatomical structure that curbs the performance of current clinical treatments for the disease. Thus, segmentation and 3D reconstruction of the LA and PVs are of crucial importance for the diagnosis and treatment of AF.

In this context, cardiac 3D Late Gadolinium Magnetic Resonance Imaging (LGE-MRI) appear as a very good tool for cardiac tissue characterization and myocardial fibrosis detection. In fact, these images have been proofed as reliable predictors of catheter ablation success, which is often the chosen treatment for AF patients.

Several manual and semi-automatic segmentation tools from LGE-MRI scans are currently in use, but these are very time-consuming and highly prone to errors, hence the need for an automatic segmentation approach.

With the rise of deep learning and convolutional neural networks, a number of automatic schemes are being developed. In this project, we evaluate a model that has been developed at the Hospital Clínic de Barcelona for obtaining an automatic segmentation of the LA using a deep learning architecture. Concretely, we tested this model with an independent set of images from another MRI vendor, and we obtained a set of quantitative and qualitative measures to validate the results.

For the pursuit of our aims, this work begins with the state-of-the-art for LA segmentation of LGE-MRI scans and with a market analysis of the field. We then present our proposed solution together with the obtained results and the corresponding conclusions.

TABLE OF CONTENTS

1. INTRODUCTION	7
1.1. OBJECTIVES.....	7
1.2. MOTIVATION	8
1.3. SCOPE AND SPAN.....	9
2. BACKGROUND	10
2.1. ATRIAL FIBRILLATION	10
2.2. ANATOMY OF THE LEFT ATRIUM:.....	11
2.3. MAGNETIC RESONANCE IMAGING	12
2.4. ARTIFICIAL INTELLIGENCE AND DEEP LEARNING	13
2.4.1. Deep Learning	13
2.4.1.1. Artificial Neural Networks	13
2.4.1.2. Training Artificial Neural Networks	14
2.4.1.3. Evaluating Artificial Neural Networks.....	16
2.5. CARDIAC IMAGE SEGMENTATION	16
2.5.1. Manual and semi-automatic segmentation schemes	17
2.5.2. Automatic segmentation schemes: The rise of deep learning.....	17
2.6. THE CONTEXT OF THE PROJECT	18
3. MARKET ANALYSIS.....	20
3.1. DEMOGRAPHICS AND MARKET SEGMENTATION.....	20
3.2. TARGET MARKET	21
3.3. MARKET NEED	22
3.4. COMPETITION.....	22
3.4.1. Products that have reached the market	22
3.4.2. Potential products that could reach the market in the future.....	23
3.5. BARRIERS TO ENTRY	23
4. CONCEPTION ENGINEERING	24
4.1. POSSIBLE SEGMENTATION MODELS.....	24
4.1.1. Convolutional Neural Networks and U-Nets	24
4.1.2. Existing models and critical choices within the models	26
4.1.2.1. Class imbalance	26
4.1.2.2. Shape variability	26
4.1.2.3. 3D vs. 2D	27
4.1.2.4. Loss functions	27
4.2. PROPOSED SEGMENTATION MODEL	28
4.3. POSSIBLE EVALUATION SOLUTIONS	28
4.4. PROPOSED EVALUATION SOLUTION.....	29
5. DETAILED ENGINEERING	30
5.1. HCB U-NET DATASET	30
5.1.1. HCB dataset.....	30
5.1.2. 2018 Atrial Segmentation Challenge dataset.....	30

5.1.3.	Dataset summary	30
5.1.4.	Data formats	31
5.1.5.	Dataset Pre-processing	31
5.2.	THE HCB U-NET MODEL.....	31
5.2.1.	The segmentation model	32
5.2.1.1.	The encoder-decoder paths.....	32
5.2.1.2.	Group normalization.....	32
5.2.1.3.	Classification layer	33
5.2.1.4.	Modular U-Net.....	33
5.2.2.	The patch-based learning strategy.....	33
5.2.3.	The loss functions.....	33
5.2.4.	Post-processing	34
5.2.5.	Evaluation metrics.....	34
5.3.	EVALUATION DATASET	34
5.3.1.	GE Dataset.....	34
5.3.2.	Image pre-processing.....	34
5.4.	SEGMENTATION RESULTS	35
5.4.1.	Outliers.....	35
5.4.2.	Inliers and the punctuation criteria.....	37
5.4.3.	Results	39
5.4.4.	Identified problems.....	41
5.4.5.	Overcoming the inaccurate PV delineation.....	41
5.4.5.1.	Loss function.....	41
5.4.5.2.	Average volume threshold.....	42
5.4.5.3.	Lack of context learning.....	43
5.4.6.	Overcoming the inappropriate LA centering crop.....	44
5.4.6.1.	Multi-CNN for dynamic LA centering crop.....	44
5.4.6.2.	Increasing the security margin	44
5.4.6.3.	Localizing the descending aorta	44
6.	EXECUTION CRONOGRAM AND FINANCIAL VIABILITY	46
6.1.	WORK BREAKDOWN STRUCTURE.....	46
6.2.	PRECEDENCE ANALYSIS.....	47
6.3.	GANTT CHART.....	48
6.4.	FINANCIAL VIABILITY	48
7.	TECHNICAL VIABILITY.....	50
7.1.	STRENGTHS.....	50
7.2.	WEAKNESSES	50
7.3.	OPPORTUNITIES.....	51
7.4.	THREATS	51
8.	REGULATIONS AND LEGAL ASPECTS	52
8.1.	EARLY STAGE AND CONSIDERATIONS.....	53
8.2.	DESIGN AND DEVELOPMENT	53
8.3.	REGULATORY SUBMISSION	54
8.4.	POST-MARKET SURVEILLANCE	54
9.	CONCLUSIONS.....	55

ANNEX 1..... 61
ANNEX 2..... 66

1. INTRODUCTION

It is well known that the average global life expectancy is increasing and that, consequently, heart diseases are earning their position as the 21st century epidemic - putting aside Covid-19. As a matter of fact, in the last two decades, atrial fibrillation (AF), the most prevalent arrhythmia, has become one of the most important public health issues and a considerable cause of health care expenditure in western countries [1].

Even though AF is associated with significant morbidity and mortality, there is still a substantial lack of basic understanding of the left atrium (LA) and pulmonary veins (PVs) anatomical structure that curbs the performance of current clinical treatments for AF [2]. On this matter, antiarrhythmic medications have shown little success, making room for catheter ablation, which presents better efficacy and outcomes. However, published literature repeatedly suggests a need for better patient selection criteria when it comes to catheter ablation techniques, given the potential complications that may arise in those patients that are not suitable for the procedure [3].

That is when the need for a better understanding of the left atrial anatomical structure comes to play. Indeed, the assessment of LA diameter and LA volume, as well as the assessment of the PVs anatomy have been set as a valid predictor of AF recurrence following radiofrequency catheter ablation. In a nutshell, the analysis and quantification of the LA and PVs of candidate patients for catheter ablation has become an essential step for the prognosis of AF in these subjects [4].

In order to assess and quantify this, three-dimensional Late Gadolinium Magnetic Resonance Imaging (LGE-MRI) has resulted to be a very efficient choice.

Parallel to the innovations going on in the world of cardiac imaging, the never-ending world of Artificial Intelligence (AI) finds itself in a turning point with the relatively recent explosion of Deep Learning (DL). As a matter of fact, one of the horizons that DL has opened is the effective segmentation of medical images, amongst which we find cardiac LGE-MR images.

Image segmentation consists in partitioning an image into meaningful structures by assigning a label to every pixel in it, such that pixels with the same label share certain characteristics. Indeed, the segmentation of the LA from cardiac MRIs is used in order to analyze the structure of the LA in detail. This job has traditionally been done either manually or with semi-automatic tools, which are very time consuming and error-prone.

In this work we use an existing DL segmentation model to assess its performance when tested with a type of images that differ from the training dataset type, to ultimately evaluate how generalizable this method would be in a specific hospital environment.

1.1. OBJECTIVES

The aim of this project is twofold.

On the one hand, the big picture shows the healthcare necessity of improving the treatment outcome of AF patients, especially of those who undergo catheter ablation. As aforementioned, LGE-MRI plays an important role in the treatment and prognosis of AF patients. This is because knowing the LA anatomical structure of the patient prior to any AF treatment can help determine the best treatment approach and, indeed, 3D LGE-MRI sequences enable the use of quantitative

post-processing tools, i.e. segmentation tools, to analyze potential treatment complications. Hence, the need for an accurate image segmentation tool. There is also an urge of saving healthcare professionals time and effort in the image segmentation process, more often than not, cumbersome and imprecise, so that they can invest their energies in other more highly specialized and/or human tasks. Hence, the need for an automatic image segmentation tool. Thus, the first objective of this thesis is to work with an automatic image segmentation tool for AF patients, to ultimately improve their treatment procedure and outcomes.

On the other hand, there is the more specific need from the Arrhythmias Section at Hospital Clínic de Barcelona (HCB). That department has recently acquired a GE Healthcare MRI machine, with which physicians obtain most of the LGE-MRI scans for AF ablation procedures. At the same time, the researcher Jana Reventós, has recently developed a DL segmentation algorithm for LGE-MRI images that has mainly been trained with 3D scans from the Siemens MRI machine also at HCB. From now on, this model will be referred to as the “HCB U-Net model” or simply the “HCB U-Net”.¹ Therefore, the main objective of this project is to assess the viability of using the HCB U-Net model with images from the GE Healthcare MRI machine, so that clinicians from the Arrhythmias Section at HCB are closer to automatically segmenting these images in pre-ablation processes.

Overall, the objectives can be summarized as:

- A. Understanding the functioning of an existing automatic image segmentation model based on deep learning. More specifically, the model, known as HCB U-Net, aims to segment the left atrium from three-dimensional late gadolinium enhancement magnetic resonance images.
- B. Assessing the viability of using the HCB U-Net model when the testing is replicated with an independent sample from GE Healthcare.

To achieve these objectives, this project begins with the state-of-the-art for LA segmentation of LGE-MRI scans and with a market analysis of the field. We then present our proposed solution together with the obtained results and the corresponding conclusions.

1.2. MOTIVATION

The high prevalence and severity of AF, as well as the room for improvement in the present segmentation techniques are the two main factors that motivate this project.

The fact that J.Reventós’s research project was available for our further work, was also an extremely important trigger to carry out this thesis.

Apart from the necessity and the good timing, this project is also driven by a personal motivation and keenness towards the biomedical imaging and the AI fields. For me, it has been a unique opportunity to take part in such an innovative and challenging project.

¹ The Final Master Thesis is titled: “Automatic Segmentation of the Left Atrium from Late Gadolinium Enhancement 3D Magnetic Resonance Images” and it was delivered on the 19th of April 2021. The code is available on github [57].

Last but not least, the fact that this endeavor goes hand in hand with a renowned bioengineering company, as it is GE Healthcare, and with the prestigious Hospital Clínic de Barcelona, has also been an important push. Not only for the reputation that these institutions entail, but also because of the feasibility that this has added to the project.

1.3. SCOPE AND SPAN

Before getting too deep into the project's details, it is very important to have a clear idea of what aspects limit our work and which others make it possible. The objectives of this thesis have already been described. Nonetheless, there are a couple of tangible milestones that are worth bearing in mind when assessing the work's span.

In the first place, as mentioned above, we aim to gauge the efficacy of the HCB U-Net model when used with images from GE Healthcare. Such a quantifiable outcome gives a pretty good idea of the project's scope.

On a secondary level, we would like to lay out the state-of-the-art methods for LA segmentation of LGE-MRI scans, to understand their limits and to see how far our model is from the rest of the existing models. Having that, we intend to study the viability of bringing HCB U-Net model into the market, given the real value that it could bring both to patients and professionals, as well as the unlimited potential and continuity that it would add to the project.

At the same time, one should not forget some of the roadblocks that can typically be found when undertaking an endeavor like this one. These complications can be summed up into two main topics: the DL technology, and the timeframe.

On the one hand, we must talk about the limitations of Deep Learning. With digitalization, in the last 10-20 years, our society has gone from having insignificant amounts of data to having huge amounts, from which large neural networks (NN), the schemes that DL algorithms follow, fuel. Not only scale of data is important, but also scale of computation has determined our ability to train large NNs. On top of this, tremendous algorithmic innovations have increasingly been seen as well, making NNs run much faster. In other words, the three factors that drive DL progress are: data, computation and algorithmic improvements.

The other side to this coin, however, is not as positive. All the aforementioned improvements, also account for demanding requirements when it comes to implementing a DL model. That is why, three of the limitations of this project are: availability of high-quality LGE-MR images, availability of strong computational power, and having a good developer at our disposal.

On the other hand, as it happens in most research thesis, time is a really valuable resource, mainly because of the existence of deadlines. In this case, we have had to adhere to two time limits: the academic deadline and the one that arises from the technological obsolescence threat. The latter, especially in the case of DL, it is very easy to overdue since the technology is evolving at an extremely accelerated pace.

2. BACKGROUND

2.1. ATRIAL FIBRILLATION

AF is the most prevalent form of sustained cardiac rhythm disturbance in clinical practice, affecting more severely the aged population. In fact, it has been referred to AF as one of the largest epidemics and public health challenges to come, especially since global life expectancy is increasing. Evidence suggests that the incidence and prevalence of AF have raised over the last 20 years and are expected to continue rising for 30 more years. If no efforts are dedicated to stopping this epidemic, almost 18 million people will suffer this condition in Europe by 2060 and 6-12 million people will suffer it in the US by 2050 [5]. Moreover, AF carries a noticeable economic burden. Concerning that matter, Kim et al.[6] demonstrated in 2011 that individuals with AF in the US had 73% higher medical costs compared with matched control subjects.

AF is a functional disorder and, according to recent studies, it is also strongly related with a structural factor.

From a functional point of view, AF could be defined as a specific type of atrial tachycardia in which the atria beat very rapidly and chaotically, i.e. they quiver. This condition has two main consequences. First, the fast pace can be transmitted into the ventricles, making them beat in an inefficient way and cause hemodynamic problems on the patient. Second, the fact that the LA is no longer capable of beating efficiently, causes blood to stand within the chamber and ultimately increase the risk of stroke.

Regarding structure, the presence of anatomical atrial sites which affect AF initiation (ectopic foci) and maintenance has been observed with modern mapping and imaging techniques. These sites are mainly the pulmonary veins (PVs) and the surrounding left atrial posterior wall.

AF is often described as a progressive rhythm disorder, since there exist three stages of the condition, moving gradually from paroxysmal to permanent. Paroxysmal AF is the mildest type, and it usually comes on suddenly. It may occur occasionally or frequently, is mostly asymptomatic, and it terminates spontaneously within days. Moving on towards a more severe level, persistent AF also comes on suddenly and is the type that continues requiring electrical or pharmacological cardioversion in order to terminate. Finally, permanent AF is characterized by an unsuccessful termination and a joint decision between patient and clinician not to pursue control rhythm treatment [7].

Even though AF is not considered to be a deadly disease *per se*, the problem with AF is its association with significant morbidity and mortality, causing 20% of strokes in people aged over 60 years [2]. To solve this, catheter ablation procedures can restore a sinus rhythm in these patients, since they can target the specific atrial sites that are causing the disturbance [8].

As seen previously, the classification of AF that is currently in use is merely based on the duration of the pathology (paroxysmal, persistent and permanent AF), and the class in which the patient falls has an important weight on the decision-making process of the treatment. To this effect, this patient classification method is ambiguous and may lead to treatment errors, especially taking into account the fact that the disease is represented both by a functional and a structural disorder.

Therefore, these patients should also be categorized in terms of underlying etiology, risk factors, and mechanisms, so that the treatment of choice is more suitable for them [8].

As a matter of fact and very importantly, it has been claimed that a high quantity of left atrial wall fibrotic tissue in the LA may account for a poor outcome of the ablation therapy [3]. Furthermore, it is known that patients with a larger LA volume are more likely to have AF recurrence after a catheter ablation procedure [4].

Hence, being able to carry out a detailed analysis of the LA structure before conducting a catheter ablation procedure on a candidate patient, would most certainly improve efficacy and outcomes of this treatment, given that the patients' selection would be more accurate and well acquainted. To this end, the LGE-MRI technique comes in very handy.

2.2. ANATOMY OF THE LEFT ATRIUM:

The role of the LA in the development of AF has been known for some time; nonetheless, recent evidence has shown that the pulmonary veins also play a critical role in the origin and treatment of AF. The PVs and LA have many anatomic similarities due to their primitive common PV origin. That is why, around 90% of PVs contain atrial myocardium, which is electrically active, and is frequently discontinuous and fibrotic. In fact, all of the patients with a history of AF have myocardium in the PVs and an increased rate of structural abnormalities, which may result in the abnormal electrical activation that causes AF. These sites from which the AF is induced are known as ectopic foci, and is where the radiofrequency ablation is targeted. Put simply, knowing the anatomy of the PVs and LA of every patient is very important in order to carry out an efficient ablation procedure [9].

The LA is located posteriorly in the chest, adjacent to the descending aorta. (Fig. 2) Four PVs usually enter the LA (Fig.1): right superior (RSPV), right inferior (RIPV), left superior (LSPV), and left inferior (LIPV). Each of the veins is directed laterally, with the superior veins directed anteriorly, and the inferior veins directed posteriorly. The LSPV has a more cranial angulation and is usually not easily seen from the axial plane.

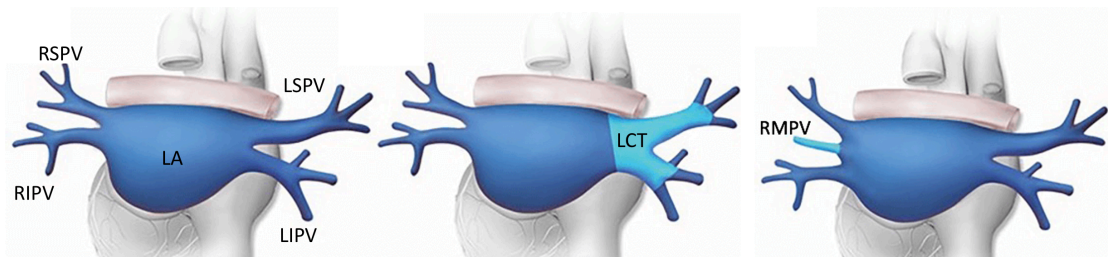


Figure 1. Anteroposterior (AP) views of the LA and PVs. Second and third show typical anatomical variations.

The anatomy of the LA does not vary much amongst patients, however, non-pathologic variations in the PVs anatomy (Fig.1) happen in approximately 40% of patients [9]. The most common ones are the presence of a single left common PV, also known as left common trunk (instead of LSPV and LIPV, they have LCT), and an accessory right middle PV (RMPV). These variations occur due to more or less incorporation of the primitive common PV into the body of the LA.

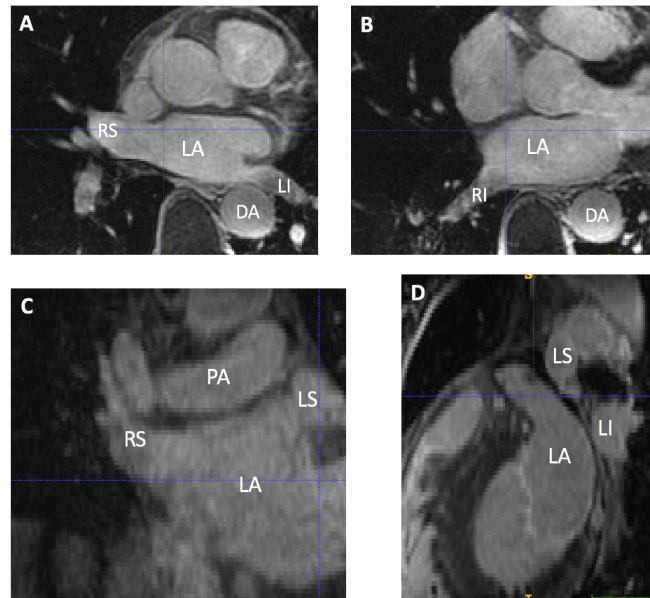


Figure 2. 3D LGE-MRI showing the anatomy of the LA and PVs. These images show the four pulmonary veins, along with the left atrium (LA), descending aorta (DA), and pulmonary artery (PA). The right inferior (RI), right superior (RS), and left inferior (LI) pulmonary veins are shown in the axial plane (A and B). The left superior (LS) and right superior pulmonary veins, and PA are shown in the coronal plane from the anterior-posterior orientation (C). The LS and LI pulmonary veins are shown in the sagittal plane (D, anterior to the left).

2.3. MAGNETIC RESONANCE IMAGING

Magnetic Resonance Imaging (MRI) is a non-invasive and non-ionizing medical imaging technique that uses strong magnetic fields, magnetic field gradients, and radio waves to generate body images. In broad terms, the electromagnetic forces in MRI target hydrogen protons in the tissues and then, each tissue returns to its equilibrium state after excitation by the independent relaxation processes of T1 (magnetization in the same direction as the static magnetic field) and T2 (transverse to the static magnetic field). While T1-weighted images are useful for defining morphological features and for post-contrast imaging, amongst other uses, the T2 modality is used to detect edema, inflammation, and other vascular events.

Cardiac MRI is of great interest when combined with contrast agents, such as gadolinium (Gd). Gd is a paramagnetic element that reduces the T1 relaxation time; hence its application in T1-weighted sequences. It consists of an intravenous injection of a 0.1-0.2 mmol/kg bolus of contrast and, after 10-20 minutes from the injection, when the substance has reached the desired body tissue, the myocardium in this case, the image acquisition begins. The main challenge found when using Gd contrast agents is finding the adequate inversion time that nulls the signal from normal myocardium, so that the myocardial cavity, and most importantly, scar tissue, exhibit higher intensity [10].

In essence, the Gd contrast agent in cardiac MRI produces Late Gadolinium Enhancement MRI (LGE-MRI) scans, which are a very good tool for cardiac tissue characterization, in particular, the assessment of regional scar formation and myocardial fibrosis. This, as mentioned above, turns

out to be a very valuable piece of information when it comes to assessing an AF patient's prognosis before carrying out catheter ablation procedures.

2.4. ARTIFICIAL INTELLIGENCE AND DEEP LEARNING

Artificial Intelligence (AI) is a discipline within the computer science field that aims to simulate human intelligence in machines. Machine Learning (ML) is a subset of AI that includes algorithms that parse data, learn from it, and then apply that knowledge to make informed decisions. ML models are capable of performing a function with the data given to it and get better progressively over time.

In general terms, the steps that should be followed when building a ML model are:

1. **Data collection**
2. **Feature extraction:** It consists in manually choosing the characteristics that make the data recognizable in order for the machine to learn these features and create internal patterns.
3. **Data preparation:** This step aims to recognize and minimize any potential biases in our datasets.
4. **Model choice:** Each model can serve different purposes and we should always look for the one that fits our necessity best.
5. **Training:** The bulk of the learning process is done in this stage. It consists in passing the data several times through the model, so that the algorithm progressively tunes the parameters according to the desired output.
6. **Evaluation and testing:** Once the model is trained, it needs to be tested with a different dataset than the one used in training to see if it operates well in real world situations.
7. **Hyperparameter tuning:** If the evaluation is successful, the hyperparameter tuning step tries to improve upon the positive results achieved during the evaluation and testing steps.

2.4.1. Deep Learning

Deep Learning (DL) is a subset of ML that is capable of automatically extracting the features from the input data. So, instead of manually selecting the characteristic traits from each piece of data, as ML models do, DL models create internal associations that place a degree of importance to each part of the data. A commonly used architecture within DL are the Artificial Neural Networks (ANNs), which represent a collection of connected and tunable computational units, called artificial neurons, organized in a layered structure of ANNs.

DL has evolved hand-in-hand with the digital era, which has brought about huge amounts of data that traditional algorithms cannot take full advantage of. In addition, DL has benefitted from advanced computer hardware, such as graphical processing units (GPUs) and tensor processing units (TPUs), making it possible to train large NNs.

2.4.1.1. Artificial Neural Networks

ANNs are computational architectures inspired by the way biological nervous systems, such as the brain, process information [11]. The ANN consists of multiple layers of simple processing elements called neurons, which perform two main functions: collection of inputs and generation of an output.

In neural networks, each node (neuron) performs some simple non-linear computations, and each connection conveys a signal from one node to another, labeled by a number called the “weight”

that indicates the extent to which a signal is amplified or diminished by the connection. In spite of the relative simplicity of the individual units, systems containing many neurons can generate complex and powerful behaviors.

Each functional unit, i.e. neuron, contains three basic components of importance (Fig. 3). First, the inputs of the neuron are multiplied by the respective weight. Second, an adder sums up all the inputs modified by their weights and may add a bias term. This is referred to as linear combination of inputs. Third, an activation function controls the amplitude of the output of the neuron, which usually ranges between 0 and 1.

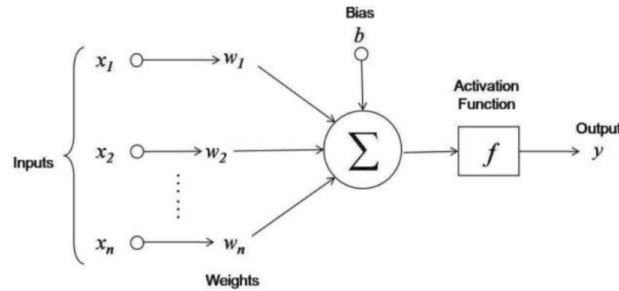


Figure 3. Diagram of the parts of an artificial neuron.

In most cases, the connections within ANNs are allowed from a node in layer i only to nodes in layer $i+1$, following a feed forward pipeline (Fig. 4). Conceptually, nodes in successively higher layers are capable of abstracting successively higher level features than the preceding layers.

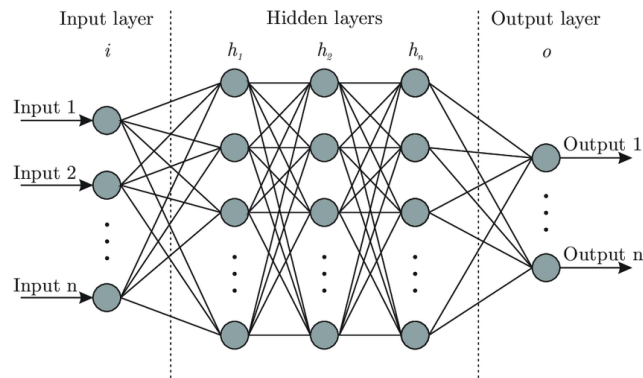


Figure 4. Typical ANN architecture.

There is always a learning method associated to each ANN, which consists in modifying the initial parameters (weights and bias) of the neural network through a continuous process of simulation by the environment in which the network is embedded. To do so, most ANNs learn in a supervised way, in which the network is trained by providing it with input and matching output patterns, known as ground truth. This means that we provide the algorithm with examples of the inputs and outputs that we want the network to compute, and then the error is calculated. This error is calculated with the loss function, which reflects the difference between actual and expected results.

So, first, signals are sent forward, and then the error is propagated backwards so that weights and bias can be modified to ultimately make the loss function decrease.

2.4.1.2. Training Artificial Neural Networks

After defining ANNs from a conceptual point of view, we will look at them from a more mathematical perspective. First, some important concepts will be defined:

- **Activation function:** It is the function that controls the amplitude of the neuron's output. Thus, the output of the neuron is a function of the linear combination of inputs:

$$y = f(z) = f\left(\sum_i x_i w_i\right) = f(W^T x + b) \quad (1)$$

Where f is the activation function. The most used activation functions are the Sigmoid and the Rectified Linear Units (ReLU) functions (Fig. 5):

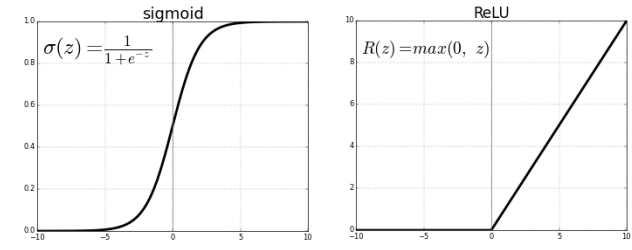


Figure 5. Graphical representation of the Sigmoid and ReLU functions.

While the sigmoid function follows the typical logistic regression morphology, in practice, the ReLU function shows better convergence performance.

- **Normalization function:** In the last layers, instead of using the normal activation function, the Softmax function is used in order to make sure that the components will add up to 1, so that they can be interpreted as probabilities.

$$\sigma(z) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (2)$$

- **Loss function:** It is the function responsible for measuring the error between the real and the expected outputs. The most commonly used loss function for the task of image segmentation is a pixel-wise cross entropy loss, which examines each pixel individually, comparing the class predictions (depth-wise pixel vector) to the target vector. Hence, this loss asserts equal learning to each pixel in the image. This can be a problem if the classes from the input data have unbalanced representation in the image, as training can be dominated by the most prevalent class.
- **Gradient descent:** It is the tool responsible for back propagation. It measures the correctness of the output after data has been fed forward and also of reassigning new values to the parameters so that the loss function is minimized.

Using these definitions, we list the steps that need to be followed in order to train an ANN:

1. **Data processing:** The input data may need to be preprocessed in order to ensure that all samples are comparable. That includes steps like normalization or resizing.
2. **Feed forward:** This step consists in generating an output with the model. The first time that an image is fed to the algorithm, the obtained output will most likely not resemble the ground truth (expected output), so the feed forward process will have to be repeated until an accurate segmentation is obtained. Instead of using a single data point in this step, several can be used, and this is known as a batch. Larger batches make the training more reliable and stable but are also more memory-consuming.
3. **Loss function:** it tells how far the obtained output is from the expected one.

4. **Back propagation:** It activates gradient descent, which is an iterative algorithm that uses the training dataset to update the model in a way that the loss function is minimized at the same time that weights and bias are optimized.
5. Repeat steps 2 to 4 until the loss function is minimized or the epochs reach the set limit. Each epoch represents one time that the data goes through the forward-backward process.

2.4.1.3. Evaluating Artificial Neural Networks

The last step when implementing a DL model is the evaluation phase, that aims to quantitatively assess the model's performance on unseen data. It is very important to know if the model is just memorizing the data it is fed with, or if it is actually learning relevant generalizable features from that data.

Often when building a DL model, the dataset is divided into these three subsets:

- **The training set:** it is the one used to build predictive models and change weights inside the network.
- **Validation set:** it is used to evaluate the model during the training phase, so that hyperparameters can be tuned and the best performing model can be selected. It is not present in all networks.
- **Test set:** it is the unseen data used during the evaluation or testing phase. It is used to assess the likely future performance of the model. By comparing the results of this set with the ones from the training set, one can see if the model is overfitting the data. Hence, the importance of the test set.

There are several metrics that are used to assess the performance of the test set. The most common metrics used in segmentation problems are the Dice Similarity Coefficient (DSC) or Dice Score, and the average Hausdorff distance.

The DSC is essentially a measure of overlap between two samples ranging between 0 and 1, where 1 denotes complete and perfect overlap. For binary data, the DSC can be calculated as:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (3)$$

Where $|A \cap B|$ represents the common pixel labels between the obtained and expected outputs.

The average Hausdorff distance is a performance measure to calculate the distance between two point sets. It can be understood as the greatest of all distances from a point in one set to the closest point in the other set. Therefore the Hausdorff distance will be equal to zero when the ground truth and the predicted segmentations are exactly the same.

2.5. CARDIAC IMAGE SEGMENTATION

Image recognition processes for medical applications include classification and segmentation (Fig. 6). In image classification, a single label is assigned to the whole image, whereas in image segmentation a label is assigned to each pixel (in 2D images) or voxel (in 3D images). For example, we would use image classification if we were trying to classify the stage of a tumor. However, image segmentation would come in handy if our objective were to identify the shape of that tumor.

Hence, we could describe image segmentation as the process to extract regions of interest from an image.

Within the image segmentation field, one can identify the following types: semantic segmentation and instance segmentation (Fig. 6). The first type is a technique that detects, for each pixel or voxel, the object category it belongs to and it treats multiple objects of the same class as a single entity. The second type treats multiple objects of the same class as distinct individual instances.

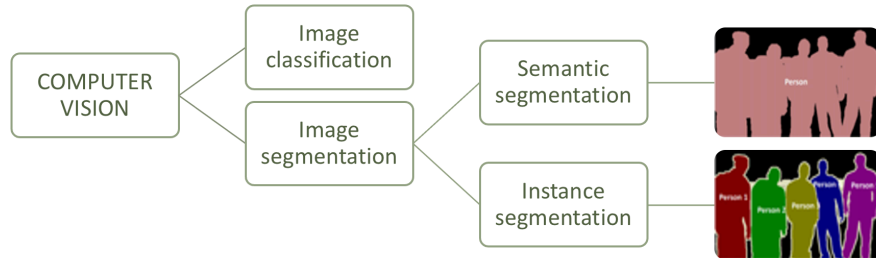


Figure 6. Subsets of the Computer Vision field.

Therefore, we can apply semantic segmentation methods to the LGE-MRI scans in order to obtain information about the LA volume and the PVs or the amount of fibrotic tissue in the LA. In this project, we will be working with the HCB U-Net, which is capable of outputting a semantic segmentation mask of the LA from LGE-MRI scans from Hospital Clínic.

2.5.1. Manual and semi-automatic segmentation schemes

Image segmentation techniques have evolved from the most rudimentary ones, where the physician or expert himself draws the mask on top of the image; to cutting-edge fully-automated tools, in which nearly no human decisions are involved; passing through semi-automatic methods that embrace the combination of the two aforementioned types.

Semi-automatic segmentation frameworks, have evolved tremendously in the past two decades. For the cardiac segmentation function, researchers started with thresholding and region growing methods, then moved on to neighboring and clustering methods, and finally reached more complex approaches such as active shape model, or multi-atlases schemes [12]–[17].

In recent years, however, DL has become the most widely used approach for medical image segmentation and, inclusive, for cardiac image segmentation.

2.5.2. Automatic segmentation schemes: The rise of deep learning

There is a wide spectrum of literature that exposes the evolution of image segmentation schemes that use DL, but we are going to directly dive into DL algorithms for cardiac image segmentation. In order to understand these schemes, it is important to introduce some of the basic neural network architectures which are commonly used in image segmentation.

Convolutional Neural Networks (CNNs) are the most common type of ANNs when it comes to image analysis, even though they also have applications outside of this field. Taking inspiration from the organization of the animal visual cortex, the individual neurons in the network are capable of responding to stimuli only in a restricted region of the visual field, known as the receptive field. This is then followed by a collection of such fields that overlap each other to ultimately cover the entire visual area. (Fig. 7)

In broad terms, the hidden layers within CNNs are functional layers that often contain convolutional layers, pooling layers and/or fully-connected layers. The network then outputs a fix-sized vector where the magnitude of each element depends on the goal of the model.

The main advantage of CNNs is that they can be trained to truly understand the sophistication of images without the need of large scales of connectivity and complexity. This is the outcome of applying relevant filters, which make the network able to successfully capture the spatial and temporal dependencies in an image. In addition, CNNs do not require much pre-processing compared to other image analysis schemes, since the feature extraction processes are no longer hand-engineered, but provided by automated learning.

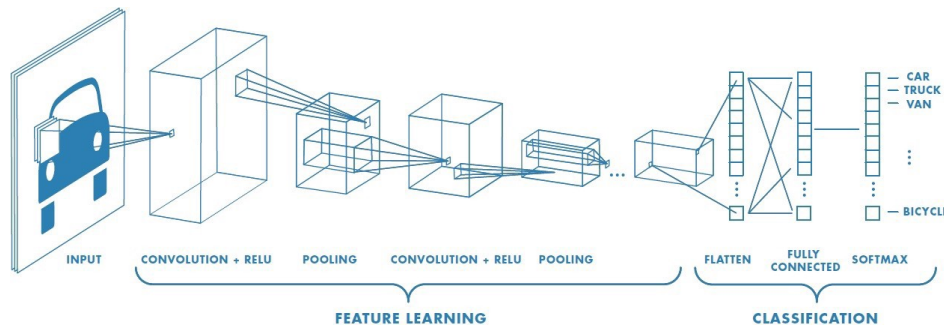


Figure 7. A typical CNN architecture.

The constant research happening in that field led in 2014 to the idea of Fully Convolutional Neural Networks (FCNs) [18], a very convenient tool to solve image semantic segmentation problems. These special type of CNNs do not have any fully connected layers and usually work within an encoder-decoder framework in a way that they can input images of an arbitrary size and produce the output with the same size.

The most notorious variant of FCNs for biomedical image segmentation is the U-Net scheme proposed by Ronneberger et al. [19] in 2015, which employs information retrieval paths between the encoder and decoder to ultimately yield a more accurate segmentation and avoid the loss of information. A year after that, Milletari et al. [20], put forward the V-Net, a U-Net modality for the segmentation of volumetric medical images (commonly called U-Net as well). U-Nets and V-Nets are being widely used in atrial segmentation problems, and the variants of this framework will be more thoroughly explained in further sections.

2.6. THE CONTEXT OF THE PROJECT

This project is originated in the Arrhythmias Section at HCB, where patients with AF are treated with catheter ablation procedures. Before the intervention, an LGE-MRI scan is obtained from each patient so that physicians can set the convenient treatment option and more or less predict the outcome of it. Currently, clinicians are manually segmenting these images with the ADAS 3D software, with which they segment a few slices and then an automatic interpolation is done to obtain the whole LA volume.

Meanwhile, the biomedical engineer J.Reventós has developed an automatic segmentation model based on a DL CNN scheme. Particularly, the network is a U-Net type network and, throughout this work, we refer to it as the HCB U-Net model.

The Arrhythmias Section, however, uses two different types of MRI machines (from two different vendors) for the obtention of the LGE-MRIs from AF patients: the 3T Siemens Medical Systems, and the 3T GE Healthcare Systems machine. The HCB U-Net model has been trained and validated with images from the Siemens MRI, and professionals within the Arrhythmias Section wonder if the model could be generalizable to the images obtained with the GE machine.

Therefore, in this project we evaluate the performance of the HCB U-Net when tested with unseen data from another MRI machine.

3. MARKET ANALYSIS

Bringing up an academic research work from its crib into a tangible product that reaches the market is never an easy task; especially in Europe, where the gap between laboratory research and market still needs some filling [21]. Though not being the primary purpose of our work, this section will analyze the viability of our research work becoming a product, so that future efforts dedicated to this project can rely on this information.

First and foremost, providing a clear idea of the hypothetical product that we are dealing with is vital (Fig. 8). The HCB U-Net model consists in a DL algorithm that inputs LGE-MRI images from the Siemens MR machine at Hospital Clínic de Barcelona and, in turn, is capable of outputting a segmented mask of the LA. In addition, in this paper we assess the precision that this same model has when inputting LGE-MRI images from the GE Healthcare MR machine present at the same hospital; and we put forward some steps that could be followed in order to make the HCB U-Net model as generalizable as possible. That said, the market analysis that will be carried out in this section revolves around a hypothetical DL model capable of building LA masks from a wide range of MRI machines that create LGE-MRI scans of the heart.

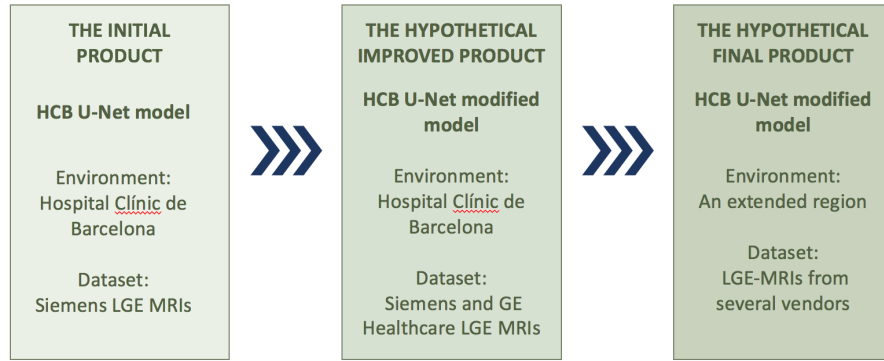


Figure 8. Evolution stages of the hypothetical product (from the HCB U-Net raw model to a hypothetical final product based on that initial algorithm).

In order to get a 360 degree view of the market that we are willing to reach, we will look into the size of that market, the existing customer segments, the competition, and the economic environment with reference to barriers to entry and regulation.

3.1. DEMOGRAPHICS AND MARKET SEGMENTATION

It is a well-known fact that the incidence and prevalence of AF are increasing globally. According to the data extracted from the Framingham Heart Study (FHS), the prevalence of AF increased 3-fold over the last 50 years, resulting in an estimated prevalence of 46.3 million individuals in 2016 suffering from this condition around the world [22]. Some of these patients can substantially diminish their symptoms with antiarrhythmic drugs (AAD) supply; however, there is still a considerable proportion of AF patients whose heart is resistant to AAD, and catheter ablation has become a well-established treatment modality for them [23].

In fact, the industry of electrophysiology ablation catheters for the treatment of AF has been projected to record an 8.5% Compound Annual Growth Rate (CAGR) globally over the period of 2020-2027 [24]. This is partly due to the ongoing R&D expenditure on that field, especially

regarding the use of robotic ablation of AF, which is expected to significantly improve the performance of the technique.

All in all, the market segment that our hypothetical product targets is clearly a robust one and even expected to grow significantly in the near future, given the global average life-expectancy increase, as well as the rise of unhealthy lifestyles in many countries, amongst other reasons. Hence, it is safe to say that the attractiveness of the market segment that we are addressing is high and it can be used in our favor in order to leverage the growth of our product.

3.2. TARGET MARKET

In this section we will get a more in depth characterization of the customer segments that we are targeting.

We have previously defined our market segment as the patients with symptomatic AF who undergo catheter ablation procedures driven by AAD resistance. First of all, we must highlight the fact that patients should not be considered within the market segment, but rather the physicians who are using the product or service. Nonetheless, the number of patients will be proportional to the number of potential doctors making use of the model.

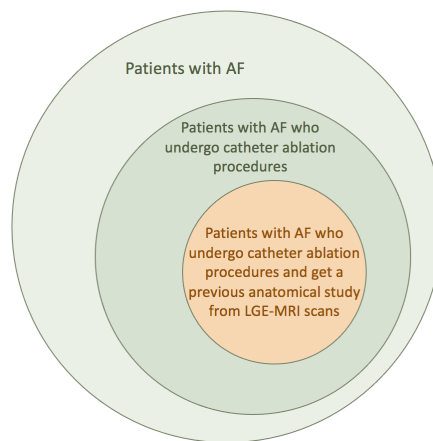


Figure 9. The customer segment that we are addressing is represented in the orange circle.

Furthermore, not all of the healthcare professionals (HCPs) who conduct catheter ablation procedures, carry out a previous anatomical analysis of the patient's heart with LGE-MRI scans. Even though the customer spectrum may seem to get more and more narrow (Fig. 9), we present a list with some of the hospitals around the world that nowadays make use of the LGE-MRI technique [25]:

- Hospital Clinic, Barcelona, Spain.
- Leipzig Heart Center, Germany
- Hospital of the University of Pennsylvania, Philadelphia, USA
- Barts Heart Centre, London, UK
- CEDIMED, Medellin, Colombia
- Hospital Universitario Virgen del Rocío, Seville, Spain
- University Hospital Bonn, Germany
- Hospital Germans Trias i Pujol, Barcelona, Spain

This list is not exclusive and should not be taken as a reference when assessing the market volume. However, it gives us a clear idea of the expansion and spread of such technique.

3.3. MARKET NEED

Here we put the spotlight on the current existing market gap and how vital it is to fill it.

A number of automatic LA segmentation methods have been designed, but most of them have been trained with very homogeneous datasets. For example, many of the existing models use the “2018 Atrial Segmentation Challenge” [26] dataset, which consists in 154 3D LGE-MRIs, all with matching parameters.

However, none of these models have yet become a market product. This is probably due to the generalization problem of these schemes, which is originated in the lack of variability within the training datasets.

The product that we put on the table aims to solve the lack of variable datasets that are used to segment the LA, especially in terms of image parameters, so that a wider range of images (from other vendors or machine models) can obtain a reliable output from these models.

3.4. COMPETITION

A very important part when assessing the viability of launching a product is analyzing which are the potential competitors in the marketplace and which are their strengths and weaknesses. After doing some research on that field, we have decided to talk about two different kinds of competitors: the ones that are already on the market, and the ones that have not yet reached it and are still in the research phase. In fact, the latter do not need to actually be willing to reach the market, they may simply be projects at an academic level. It has been done this way because, as far as we are concerned, no automatic tools are yet in use to segment the LA from LGE-MRI images.

3.4.1. Products that have reached the market

As mentioned above, there is not yet an available technology that automatically segments our region of interest from LGE-MRI scans. Nonetheless, there are a number of widespread tools that are worth mentioning.

The first software solution arises from the junction of two companies, Galgo Medical (Barcelona, Spain) [27] and Circle Cardiovascular Imaging (Calgary, Canada) [28], that developed the ADAS 3D LA tool [25]. This software enables to better visualize the fibrotic substrate in the LA for pre and peri-procedural clinical decision-making and, to do so, it first has to provide tools to define the shape of the LA. This technology, however, is not based on AI and hence is not fully automatic.

Secondly, the Cardio AI solution launched by Arterys (San Francisco, United States) [29] offers an AI-based automatic software for cardiac MRI images. Though apparently very useful, this tool only allows HCPs to automatically segment the Left Ventricle from LGE-MRI and not the LA.

Moreover, some cardiac mapping systems, like the EnSite Precision™ Cardiac Mapping System (Abbott Medical, Chicago, United States) [30], have a semi-automatic segmentation feature in them, that is also commonly used throughout medical centers.

3.4.2. Potential products that could reach the market in the future

A lot of papers have been published in the last 3 to 5 years about the automatic segmentation of the LA for AF patients. Amongst these, we highlight three of them, as they have shown to perform remarkably [31]. These are:

- Automatic 3D Atrial Segmentation from GE-MRIs Using Volumetric Fully Convolutional Network by Xia et al. [32] in 2018, with Dice Score: 93.2.
- Dilated convolutions in neural networks for left atrial segmentation in 3D gadolinium enhanced-MRI by Vesal et al. [33] in 2018, with Dice Score: 92.6.
- Pyramid network with online hard example mining for accurate left atrium segmentation by Bian et al. [34] in 2018, with Dice Score: 92.6

3.5. BARRIERS TO ENTRY

As in all markets, there exist several barriers to enter the LA segmentation market. We divide these into four categories:

- **Investment:** Getting funding is never easy. However, given that the market that we are targeting is considerably active, this part should not be an especially complicated one.
- **Technology:** Mainly the fact that the DL is advancing so fast, could render our product obsolete if no continuous research was done.
- **Regulation:** It will be seen in a further section, that the main regulation for these products in the European Union is the MDSW regulation.
- **Access to clients:** The environment we find ourselves into is a very complete combination of physiologists and engineers, giving our product a competitive advantage with respect to other competitors.

4. CONCEPTION ENGINEERING

4.1. POSSIBLE SEGMENTATION MODELS

In this section, we will discuss the possible models that could have been used to segment the LA from LGE-MRI images. Since most of them involve CNNs, and particularly U-Nets, the first part will delve into the details of these architectures.

4.1.1. Convolutional Neural Networks and U-Nets

Convolutional Neural Networks

CNNs are built on the idea that each artificial neuron within the network receives input from nearby pixels or voxels in the image, hence being capable of extracting repeating features such as borders, texture, etc.

The functional unit of a CNN is known as convolutional filter (Fig. 10). Each convolutional filter is defined by its size of kernel and stride length:

- Size of kernel: number of neurons from the previous layer that input to the next layer's neuron.
- Stride length: jumps during the convolution. When stride length is 1, then the convolution is non-strided, it is a mere mathematical convolution.

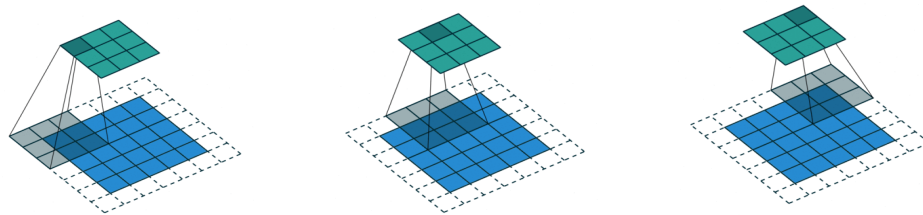


Figure 10. Convolutional process of a CNN convolutional filter.

A set of convolutional filters form a convolutional layer. In practice, filters from the same layer are run over the input image or preceding layer simultaneously, and each filter provides a different type of information. It is disregarded what each layer and filter obtain, but complexity increases as images travel deeper. For example, the most superficial layers look for basic features such as edges and landmarks, whereas deeper layers can analyze the relationship between shapes.

Pooling layers are often placed in between convolutional layers to regularize weights and make them less sensitive to noise, hence overfitting. These layers are based on the assumption that spatially close neurons should have similar outputs, and therefore can be grouped by a simple mathematical operation (maximum, mean, median, etc.) defined by a kernel of an arbitrary size. In these layers, no estimable parameters exist and that is what makes it different from a normal convolutional layer.

Fig. 11 shows a typical CNN architecture for image classification.

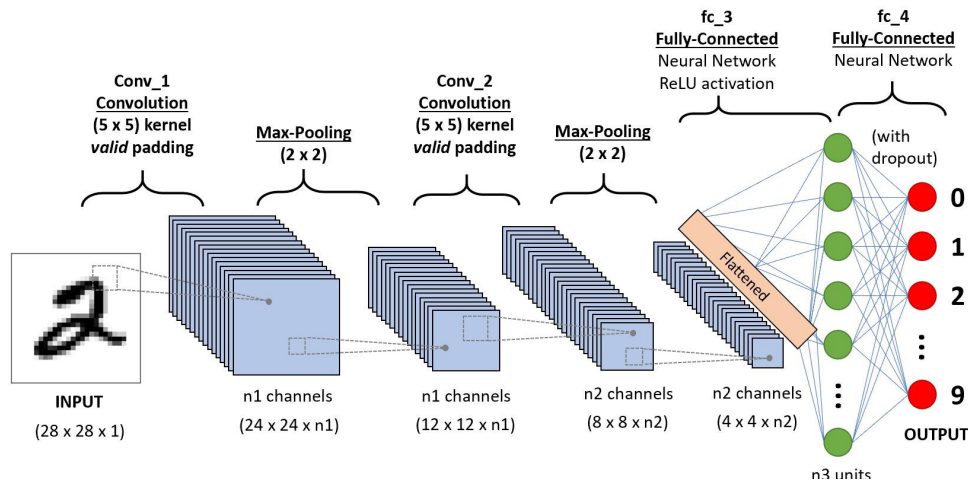


Figure 11. Typical CNN scheme for image classification.

Image segmentation problems, however, are more complex than classification problems, since they have as many classification unknown outputs as there are pixels. U-Nets are a subset of CNN that has shown very good performance in the biomedical image segmentation field.

U-Nets

The U-Net (Fig. 12) was first designed by Ronneberger et al. [19] for biomedical image segmentation and demonstrated great results on the task of cell tracking. The advantage of the U-Net is that it can achieve good results with a relatively small amount of examples. It owes its name to a U-like shape, formed by a contracting path with down-convolutional layers, and an expanding path with up-convolutional layers.

During the contracting phase feature maps get spatially smaller and then go back to the original size of the image in the expanding phase. A very important aspect about U-Nets is the fact that the up-sample representations at each step concatenate with the corresponding feature maps at the contraction path. Finally, at the last step, the architecture outputs the probability of belonging to the mask for every pixel in the image.

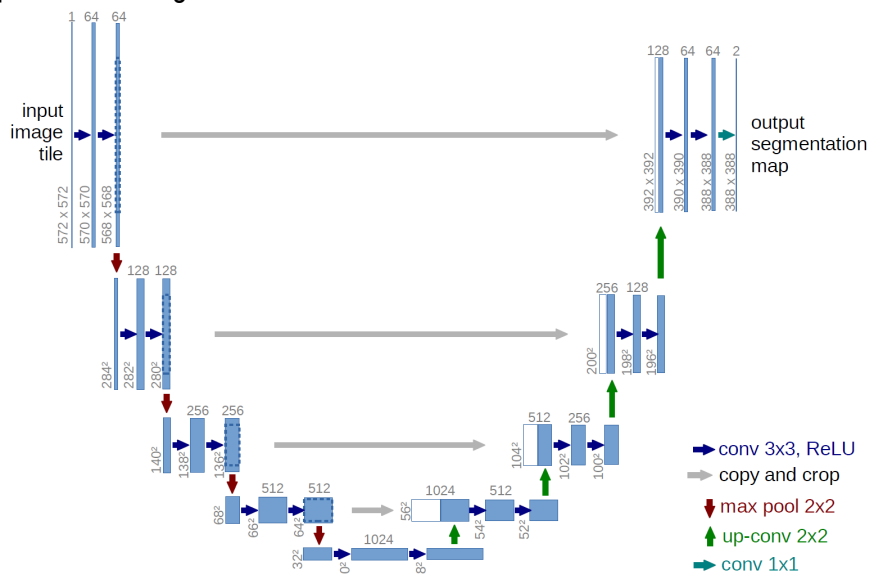


Figure 12. The Ronneberger U-Net.

4.1.2. Existing models and critical choices within the models

Achieving an automatic tool for the LA segmentation from LGE-MRI images is very challenging, especially due to the complex anatomy of the LA. One of the main encountered difficulties is class imbalance: it is known that, in average, the atrial cavity represents only a small fraction of the image volume ($\sim 0.7\%$), hence affecting the learning process given that the background is over-represented whereas atrial structures are under-represented [31]. Moreover, another problem that impairs the extraction of relevant features during the training phase and consequently decreases segmentation performance, is the inconsistency in the sizes of the atrial anatomical structures such as the PVs or the left atrial appendage seen in scans from different patients, as well as the complexity of the atrial anatomy itself. Finally, computer memory limitations and insufficient 3D data are also a problem in the implementation of these models.

In a nutshell, it could be said that class imbalance, high shape variability, and technology limitations are the three main hindrances that LA segmentation models face.

4.1.2.1. *Class imbalance*

Class imbalance occurs when the label classes from the input data have an unbalanced representation within the image. The presence of this problem in DL environments can go unnoticed, since the sensitivity, or true positive rate, may not vary even if the algorithm is missing all of the desired structure. Put simply, the model could output every pixel as background and still get a 99% sensitivity (if the atrial label accounted only for 1% of pixels). The cause of this is the fact that, during the training phase, the algorithm adjusts the weights according to the features that it considers as relevant. Therefore, if the LA structure is not seen as a relevant feature within the input image due to its small size, the model will not learn to differentiate it.

To address this, several researchers have chosen to crop the input images by means of a variety of methods, so that a bigger proportion of the LA wall is present in them. Vesal et al. [33] proposes to crop the input images from the center, using fixed coordinates. By contrast, other researchers [32], [35], [36] use a multi-CNN approach, where two consecutive networks are employed. Instead of applying fixed coordinates from the center of the image, they dynamically center the region of interest on the LA for each input. The first CNN is able to localize the LA on each image and exclusively crop background pixels, and the second network focuses on the segmentation task *per se* from the previously cropped patch.

4.1.2.2. *Shape variability*

There is a great variation across different subjects and timepoints in the shape and size of the LA. In addition, the anatomical arrangement of PVs, especially the amount of veins, do not present a consistent pattern in population [9]. That, along with the fact that the LA is really small compared to the rest of the background, explains the need for strong context information to ultimately be able to localize the LA. For example, tracing where the lung and other heart chambers are found, can substantially help segment the LA. However, this requires the network to have the ability to extract proper feature hierarchy, as well as to have a large receptive field [34].

To achieve this, Bian et al. [34] propose a pyramid pooling module that distills semantic features from multiple scales, ultimately improving network robustness against the high shape variability usually encountered in the clinical datasets. Apart from addressing the class imbalance issue, Vesal et al. also tackle the problem of shape variability by employing a 3D U-net (V-net), with dilated convolutions in the lowest level of the network, and special connections that allow the incorporation of local and global knowledge. These convolution layers use dilation rates that enlarge their receptive fields and allow the network to learn different scale features. This, however, did not result in an improved learning process nor performance.

It is worth mentioning that the pyramid pooling technique used by Bian et al., can also be applied in the context of image classification. For example, some researchers [37] have used it to extract added information from the dataset and classify the images between pre-/post-ablation.

4.1.2.3. 3D vs. 2D

The largest 3D LGE-MRI dataset nowadays belongs to the Cardiac Atlas Project [38] and contains 154 scans, which represent more than 13.500 2D images. This is a relatively small amount of data when compared to the quantity that is normally used in DL algorithms. This is one of the reasons why it is important to choose wisely between training a model with 2D or 3D images. Not only the amount of available data is important when making this decision, but also the computational cost and processing time that each type of data entails.

On the one hand, volumetric networks imply a large number of trainable parameters, leading to high memory consumption, and correspondingly more processing time. On top of that, small datasets cannot be exploited efficiently compared to 2D methods, where each volume accounts for several slices. Nonetheless, the advantage of 3D approaches lies within the fully exploitation of data dimensionality and the inter-slice continuity achieved during training, which leads to a better spatial representation and a more accurate output 3D anatomy. The real benefit of this is the feasibility for a real clinical application, where precision and reliability are always valuable assets [39]. Moreover, the current memory limitations of this method will decrease as GPU technology keeps improving, making 3D approaches even more convenient. Also, datasets are growing as digitalization takes over, and small datasets will no longer be a real problem.

2D schemes, on the other hand, can reduce the risk of overfitting since more amount of data can be inputted into the network. Moreover, they require considerably less trainable parameters, leading to lower memory requirements, as well as time. The downside, however, is that these schemes are not capable of learning good spatial context features, resulting in lower performance. In essence, the strengths that 2D methods gather, are translated into weaknesses in the volumetric field, and vice versa.

In order to have the best of both worlds, Puybareau et al. [40] propose a “pseudo-3D” method for LA segmentation, in which 2D images are color expanded into the RGB space to generate a three-channel image. Even though this approach does not provide great spatial resolution, it can be taken into consideration if resources are limited.

4.1.2.4. Loss functions

Apart from trying to solve some of the core problems that the LA segmentation carries, other researches have also developed novel methods to refine the accuracy of the segmentation.

The current standard loss functions used in DL schemes for segmentation tasks are the cross-entropy loss function or the dice loss function. These losses, however, are known as region-based losses and place more importance to volumes than to contours, potentially hindering the learning of accurate boundaries. A possible solution to this problem lies in the use of contour-oriented loss functions.

Han et al. [41] propose a boundary loss function that incorporates distance, area, and boundary information as contour features simultaneously. Also, Kervadec et al. [42] introduced a contour-based loss function with high performance in highly unbalanced datasets (like ours). The latter function is basically a distance metric on the space of contours (or shapes), not regions, hence mitigating the difficulties of region-based losses in the context of highly unbalanced segmentation problems.

Moreover, also for the purpose of decreasing the effect of a class imbalanced dataset, Sudre et al. [43] propose a Generalised Dice Loss Function. While the simple Dice Loss is a similarity score between two binary sets, the ground truth and the prediction, the Generalized Dice Loss applies a weight parameter to control the impact of the class imbalance.

4.2. PROPOSED SEGMENTATION MODEL

J.Reventós developed a 3D U-Net capable of processing local information and diminish the effects of class imbalanced data by using patches as input during training, similar to the fixed coordinate method used by Vesal et al. The network's architecture will be more thoroughly explained in the "Detailed Engineering" section.

4.3. POSSIBLE EVALUATION SOLUTIONS

Once the HCB U-Net model had been run and experimental results were obtained, it was time to move on to assessing how well this architecture works when testing 3D LGE-MRI scans from the GE Healthcare MRI machine. The problem with these images was the absence of their corresponding ground truth segmentations, which are a critical tool in order to obtain a quantitative performance evaluation.

Two approaches were considered with the aim of evaluating the model's performance with a different set of images: a quantitative approach, and a qualitative one.

All the images used during the testing phase of the HCB U-Net model carried their corresponding ground truth segmentation done by experts (expected output). This allowed J.Reventós and her team to obtain a Dice Score and the Hausdorff Distance out of the output masks (real output). That is to say, they obtained a quantitative comparison between the real and the expected outputs.

Our dataset, however, only consists in the original MRIs (inputs) and the real outputs, without their corresponding ground truths. To overcome this issue, two solutions were considered:

- A. *Generating our own ground truth set to be able to obtain a quantitative comparison between real and expected outputs, i.e. obtaining the Dice Score and Hausdorff Distance.*

Obtaining good ground truth segmentations from LGE-MRIs of the LA requires a high level of expertise. Not knowing that from the beginning, I tried to do the segmentations using the EnSite Precision™ Cardiac Mapping System Verismo segmentation tool (Abbott Medical, Chicago, Illinois, United States). Thanks to the help of some electrophysiology engineers, I learned how to use this tool but, unfortunately, this was not enough to obtain good volumes as my lack of experience within that field was obvious and the resulting segmentations had very poor quality (Fig. 13). Hence, this option was ruled out.

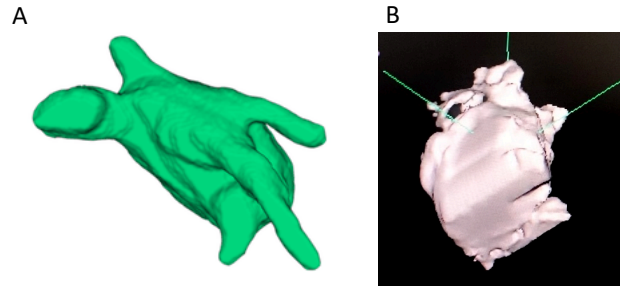


Figure 13. Comparison between (A) what an LA ground truth segmentation should look like vs. (B) the LA segmentation that was obtained.

B. Designing a qualitative evaluation method based on the clinical importance of each aspect within the LA segmentation.

A viable option for assessing the model's performance with our new images and without reference outputs, was to carry out a qualitative evaluation taking into account the important clinical aspects of the LA segmentation.

4.4. PROPOSED EVALUATION SOLUTION

The proposed solution is option B, a qualitative analysis. The most clinically transcending aspects of the segmentation of the LA for AF is the precise delineation of all of the PVs, where ectopic foci can be encountered, and the accurate segmentation of the LA volume, which is known to have an effect on the prognosis of these patients. [44]

First, some of the segmentations were classified as “outliers”, given that they were not representative of the general results; and for the remaining “inliers”, we formulated a punctuation method that will be more thoroughly explained in the next section.

5. DETAILED ENGINEERING

5.1. HCB U-NET DATASET

The model built by J.Reventós used a combined dataset source, mainly from the Department of Arrhythmias of HCB, and also from the 2018 Atrial Segmentation Challenge repository.

5.1.1. HCB dataset

On the one hand, the HCB dataset consists in 139 3D LGE-MRIs. Within these images, 107 belong to patients who underwent wide circumferential isolation of the PVs by radiofrequency in a study performed within the Hospital department [45]. The rest of the images, 32 of them, belong to patients who underwent atrial surgical ablation procedure for AF or Atypical Atrial Flutter treatment.

The first 107 images were acquired with a 3.0 Tesla MRI scanner (Siemens Medical Solutions, Erlangen, Germany), whereas the remaining 32 images were acquired with another 3.0 Tesla MRI scanner (GE Healthcare Systems, Chicago, Illinois, United States).

All 3D scans were segmented using the ADAS LA tool from the ADAS 3D image processing software (Galgo Medical SL, Barcelona, Spain). This is a semi-automatic segmentation method that requires the physiologist to manually delineate the mid-atrial myocardium of a few axial planes, so that the algorithm can interpolate these contours to the rest of the slices. The result of this ground truth segmentation is a 3D shell with wall fibrosis information.

5.1.2. 2018 Atrial Segmentation Challenge dataset

On the other hand, the 2018 Atrial Segmentation Challenge (ASCh) repository consists of 154 3D LGE-MRIs from 60 AF patients. The main data provider is The University of Utah.

The clinical images were acquired with either a 1.5 Tesla Avanto or 3.0 Tesla Verio whole-body scanner (Siemens Medical Solutions, Erlangen, Germany).

Ground truths were manually segmented in a slice-by-slice manner by experts in the field and then stacked the 2D segmentations together.

5.1.3. Dataset summary

The following is a summary table (Table 1) of the dataset characteristics and their distributions throughout the training, validation, and testing sets:

AMOUNT OF 3D IMAGES	107 (HCB)	32 (HCB)	154 (ASCH)
MRI SCANNER	3T Siemens Medical Solutions	3T GE Healthcare Systems	3T Verio and 1.5T Avanto (Siemens Medical Solutions)
SPATIAL RESOLUTION (MM)	1.25 x 1.25 x 2.5	0.78/0.625 x 0.78/0.625 x 1.2/2.5	0.625 x 0.625 x 0.625
3D LGE-MRI DIMENSIONS	320 x 320 x slices	512/640 x 512/640 x slices	576/640 x 576/640 x 88
TRAINING SAMPLES	67	27	80
VALIDATION SAMPLES	18	0	20
TESTING SAMPLES	22	5	54

Table 1. HCB U-Net dataset summary.

5.1.4. Data formats

Raw images that come out from a medical imaging device have the DICOM format. Segmentation masks, however, suffer a format transformation that is no longer standardized throughout segmentation tools. In the HCB dataset, where the segmentation is done using the ADAS 3D tool, masks are outputted in a .vtk format. It stands for Visualization ToolKit and it provides a consistent data representation scheme for a variety of dataset types. In the ASCh dataset, manual segmentation outputs .nrrd masks. The Nearly Raw Raster Data format is intended to support scientific visualization and image processing applications. These two formats, however, are not suitable for the training purpose of the *HCB U-Net* given the computation cost these entail. Hence, their transformation into Numpy arrays.

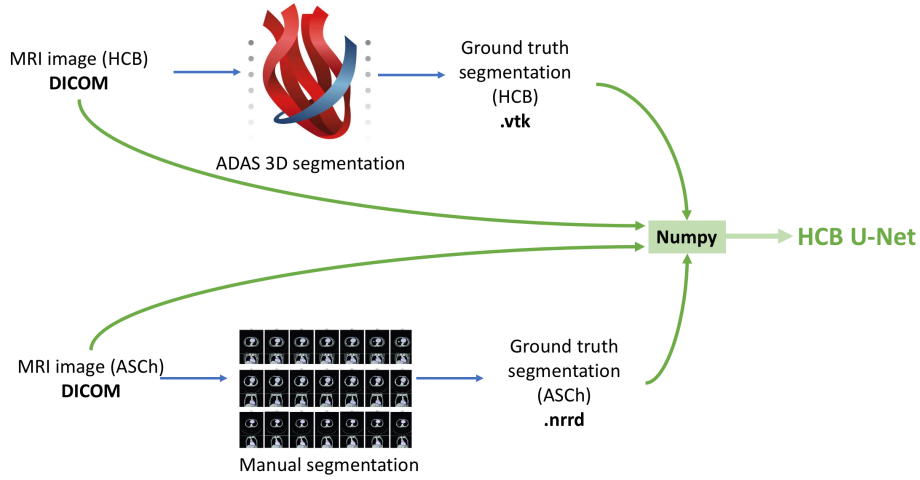


Figure 14. Data format transformation diagram.

5.1.5. Dataset Pre-processing

A set of preprocessing steps were applied to the images before they were introduced into the algorithm. These included:

1. Dimension reduction: The wide variety of LGE-MRI sources creates the need for a common dimensional scale across all datasets. This common space is characterized by a 320 x 320 x 60 dimension, and a 1.5 x 1.5 x 2.5 mm³ spatial resolution. Both these dimensions are of lower quality than the original images.
2. LA centering crop: Cropping the input images in a way that the LA structure is relatively centered, allows for the learning process to be less biased towards the background pixels.
3. 3D LGE-MRI normalization: The variety of LGE-MRI sources also generates the necessity of normalizing pixel intensities, given the wide assortment of intensities used by different vendors and scanner models.

5.2. THE HCB U-NET MODEL

Having understood some of the most commonly used schemes, the structure and characteristics of the proposed network *HCB U-Net* model will be described. First, an overview of the network and its modular structure; second, an analysis of how the network is trained in a way that it obtains local information of the heart's LA; third, a description of the various types of loss functions used during

the HCB U-Net experimentation; and fourth, a short explanation of the used post-processing and evaluation metrics.

5.2.1. The segmentation model

The essence of the model mainly lies in two ideas: the use of patches for training, and the U-Net architecture.

Regarding the U-Net architecture, this section will start describing the outer structure, i.e. the encoder-decoder paths. Then, we will delve a bit deeper into the structure and talk about the group normalization layers and the classification layers. Lastly, we will talk about the modularity of this particular U-Net.

5.2.1.1. The encoder-decoder paths

As every U-Net, the current model consists in a contracting (encoder) and an expanding (decoder) path (Fig. 15). The encoder is built with four layers, each containing two $3 \times 3 \times 3$ convolutions (unpadded), a ReLU activation, and a $2 \times 2 \times 2$ max pooling operation with stride 2.

The decoder up-samples the resulting feature map from the first phase by means of three layers, each containing one $2 \times 2 \times 2$ up-convolution with stride 2, and two $3 \times 3 \times 3$ normal convolution

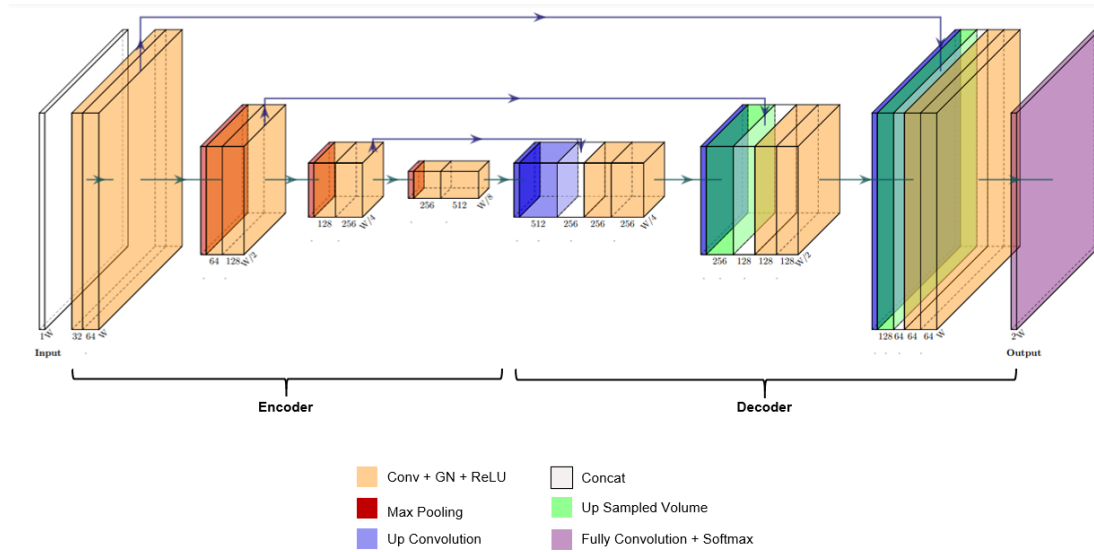


Figure 15. HCB U-Net architecture, where a encoder-decoder structure can be appreciated.

followed by a ReLU. Also, each layer is concatenated with its corresponding cropped feature map from the encoder path, providing context information crucial for the reconstruction of the feature map to the original input size.

5.2.1.2. Group normalization

The batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters. Efficient batch sizes include 32, 64, 128, or the entire dataset. For the training of the HCB U-Net, they decided to set the batch size to 1 for lower memory-consumption. Thus, here we talk about groups instead of batches.

Batch normalization is a technique for training deep networks that standardizes the inputs to a layer for each batch with the aim of stabilizing the learning process and dramatically reducing the number

of training epochs required. Since no batches were used for the training, the normalization process here is known as group normalization.

5.2.1.3. *Classification layer*

The final layer of the HCB U-Net scheme comprises a $1 \times 1 \times 1$ fully convolution that reduces the output of all channels to the number of classes, two in this case, with a final Softmax function. The final outputs are the prediction probabilities for each pixel corresponding to the foreground (LA) or background.

5.2.1.4. *Modular U-Net*

This is a very modular network, in which the layer order and the hyperparameters can be easily modified. This provides the model with versatility.

5.2.2. The patch-based learning strategy

As previously mentioned, it is very hard for models to learn in an unbiased way when the classes within the input data are imbalanced. Also, the fact that there exists a high shape variability of the LA amongst patients, makes the availability of training data very valuable.

To solve this, the HCB U-Net is trained with volume patches (small volumes within the original 3D LGE-MRI), which are used for data augmentation, and local information capture.

This patch-based strategy, however, requires a certain level of quality during the selection process. In other words, patches should not be selected in a random way, because they may not include our region of interest. To do so, J.Reventós implemented the following steps:

1. Set the patch dimensions to be lower than the whole volume.
2. Set a positive probability percentage parameter forcing the patch volume to contain a certain amount of LA, to avoid just picking background.
3. Search for complying patches.
4. Crop the patches and create a pool of 3D LGE-MRI patches and their associated ground truth.

For each data sample, the network randomly selects a single 3D LGE-MRI patch and its corresponding labelled patch. Also, it is important to bear in mind that patched images were used only for training, and not for validation and testing.

5.2.3. The loss functions

This model under consideration employs a combined loss function, which entails the weighted sum of two loss functions: the region-based and the contour-based losses.

For the region-based component, the Generalized Dice Loss is chosen, as it performs better in class imbalanced datasets, as explained in previous sections.

On the other hand, for the contour-based component, the HCB U-Net is trained using Kervadec's [42] boundary loss, which can roughly be described as the sum of the linear functions of the regional softmax probability outputs of the network.

Since the contour-based loss can be combined with a regional-based loss, the latter is used as a guide during the first epochs, and then the contour loss is incorporated in a weighted form. In fact,

contour-based loss functions are very prone to finding a solution in a local minimum or saddle point, hence requiring the presence of a region-based loss that guides it towards the real optimization. The weighted combination of the two loss functions looks as follows:

$$\mathcal{L}_{TOTAL} = (1 - \alpha)\mathcal{L}_{GDL} + \alpha\mathcal{L}_B \quad (4)$$

Where $\alpha \in [0,1)$ is the parameter that balances the two loss function types.

5.2.4. Post-processing

The LA and PVs aggregate form a continuous volume. The post-processing phase was applied to get rid of the isolated segmented regions.

5.2.5. Evaluation metrics

Last but not least, two metrics were used to evaluate the model: the DSC or Dice Score, and the Hausdorff distance. The overall metrics with optimized parameters were: 88.4% DSC with the HCB dataset, 91.0% DSC with the ASCh dataset, and accurate Hausdorff distance metrics for both datasets.

5.3. EVALUATION DATASET

5.3.1. GE Dataset

Once the model had been trained, validated, and tested, we worked for the specific purpose of this thesis with a 20 3D LGE-MRIs dataset originating from the 3T GE Healthcare Systems MRI scanner in HCB.

The obtained raw images have a spatial resolution of $0.78 \times 0.78 \times 1.2 \text{ mm}^3$, and 3D LGE-MRI dimensions of $512 \times 512 \times 112$.

In order to assess the performance of the HCB U-Net on this new independent sample, we first transformed them into the following dimensions: $1.5 \times 1.5 \times 2.5 \text{ mm}^3$ spatial resolution, and dimensions of $200 \times 130 \times 35$. This resulted in a worsening of resolution of about 20% in the x and y planes.

5.3.2. Image pre-processing

The GE Healthcare dataset was run through the model after applying a post-processing action, which consisted in cropping the images using the HCB U-Net LA centering crop dimensions. Also, we tried to input the cropped images with their original spatial resolution ($0.78 \times 0.78 \times 1.2 \text{ mm}^3$), but it was too memory-consuming. Ultimately, the masks were obtained by roughly halving the resolution to $1.5 \times 1.5 \times 2.5 \text{ mm}^3$. Moreover, it is important to highlight the fact that no patches were extracted from these image as they were meant for testing and not for training the model.

Fig. 16-A shows an example of one of the DICOM raw images, and Fig. 16-B shows the same preprocessed image. A considerable resolution and size downgrading can be appreciated.

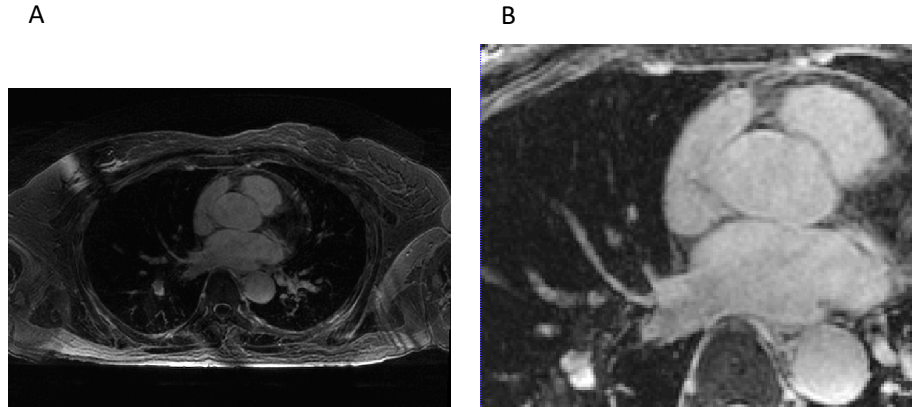


Figure 16. (A) DICOM raw image vs. (B) preprocessed image.

5.4. SEGMENTATION RESULTS

This section comments on the qualitative analysis of the viability of using the HCB U-Net for the GE Healthcare LGE-MRI dataset. First, we will overview the pre-processing of the image dataset before being fed to the model, and then we will deal with the resulting segmentations *per se*.

It is important to mention that the following analyses have been carried out by means of the ITK Snap [46] and Paraview [47] image visualization tools.

5.4.1. Outliers

This section will evaluate the resulting segmentations from the HCB U-Net. We will start by separating the outliers and commenting on those, and then we will discuss the inliers.

An outlier is a data point that differs significantly from other observations. In this sense, this definition leaves it up to the analyst to decide what will be considered abnormal. Nonetheless, before abnormal observations can be singled out, it is necessary to characterize normal observations. To do so, we have looked at all the resulting segmentations, and concluded that normal outputs must meet the two following conditions:

- Contain more than approximately 90% of the LA and PVs volume. Therefore, the LA centering crop cannot slice out more than 10% of our region of interest.
- Strongly resemble the LA and PVs aggregate without containing large parts of extra structures, nor be completely incoherent with the anatomy of the LA.

Having established this criteria, 3 out of 20 patients were categorized as outliers: patient 3, patient 9, and patient 19. The rest of them are within the inliers collective. Details of these outlier images are described below.

Patient 3

The resulting segmentation from patient 3 is considered an outlier data point because it does not comply with the first condition of normality. It must be emphasized, however, that the cause of this abnormality is not within the HCB U-Net network itself, but rather within the data pre-processing phase. The LA centering crop set for the HCB U-Net is not suitable for this specific patient.

Such a significantly inappropriate LA centering crop only happens to 1 out of our 20 patients (5% of the sampled population) and hence should not be considered a critical concern. However, as it will be exposed in the following section, a significant number within the inlier data points also suffer inappropriate LA centering crop to some extent. This will be dealt with in the inliers chapter.

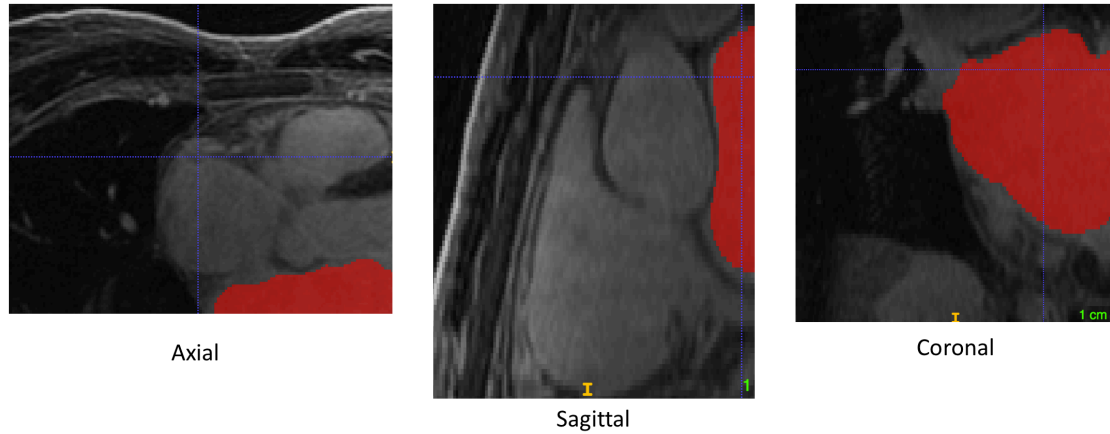


Figure 17. Axial, sagittal, and coronal planes of the Patient 3 segmentation.

Patient 9

The problem with patient 9 lies in the fact that the segmentation includes extra anatomical structures, in this case, the pulmonary artery (PA):

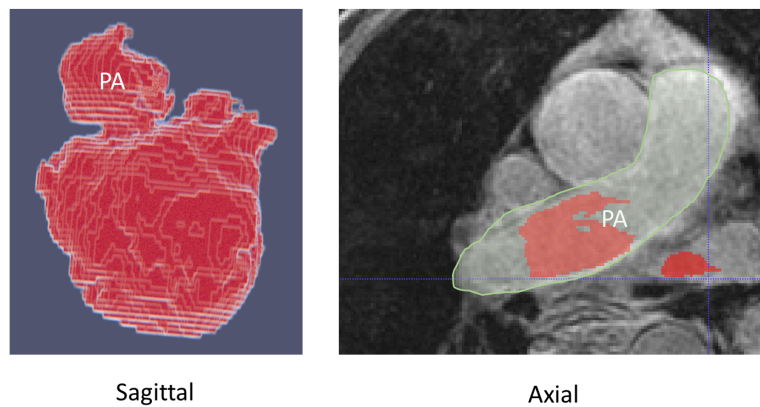


Figure 18. 3D reconstruction and axial plane (PA in green) of the Patient 9 segmentation.

The cause of this flawed segmentation is probably the poor image quality. MRI quality depends on spatial resolution, image contrast, signal-to-noise ratio (SNR), and the presence of artifacts. Spatial resolution does not differ from the rest of the images, contrast should be similar to the rest of the dataset, even though it can slightly vary amongst different patients, and there are no artifacts. However, the SNR is probably low in this image as many dots can be appreciated throughout the whole image.

The presence of noise may have led the network to segment the PA, but another important mistake in the segmentation of this patient's heart is that no PVs are delineated:

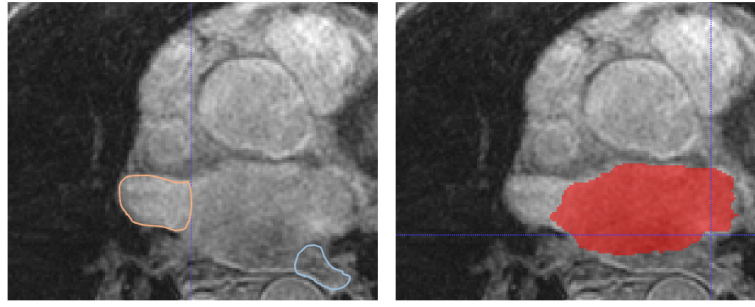


Figure 19. Axial views of the Patient 9 cropped LGE-MRI. In orange, the RSPV, in blue, the LIPV, and in red, the actual segmentation. It can be observed that none of the delineated PVs are included within the red region.

This error rises the suspicion that the model probably applies high weights to the volume measure feature. That is to say, that once the segmentation has reached a certain volume, the model stops segmenting other sites because it has learned an average volume threshold. In this case, since it “wastes” some of this volume for the PA, it is logic that the PVs are not even considered by the model.

Patient 19

This patient’s MRI scan also has a very low SNR and the network is not able to properly detect the anatomy of the LA and PVs. Moreover, it can be seen how the model still tries to reach that threshold segmentation volume mentioned previously by segmenting any remotely similar structure that it finds within the image.

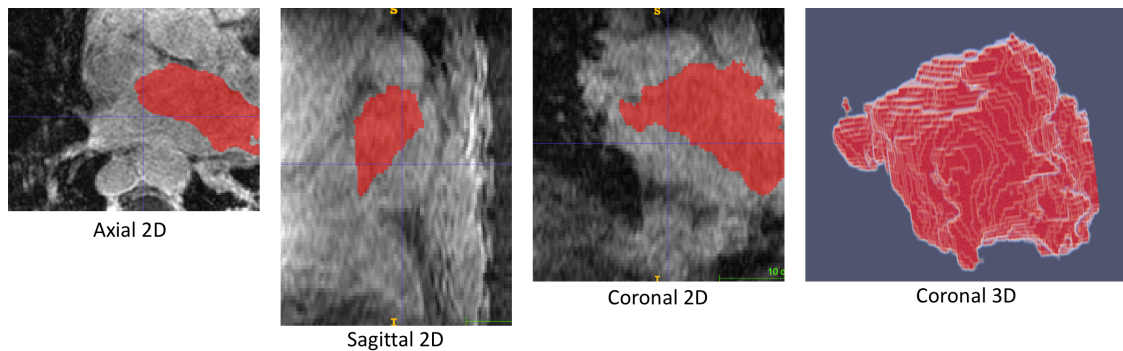


Figure 20. Axial, sagittal, coronal and 3D reconstruction seen from the coronal view of Patient 19 segmentation.

5.4.2. Inliers and the punctuation criteria

As opposed to outliers, inliers are data points that fall within the limits of normality amongst all of the observations. In this example, inliers contain at least 90% of the LA and PVs anatomical structure, and are coherently segmented.

Before determining whether a segmentation is accurate or not, an evaluation criteria should be set. With this purpose, some clinically relevant aspects that must be taken into account when evaluating the segmentations are exposed.

Researchers have identified a number of predictors for AF radiofrequency ablation success, including comorbidities, the type of AF, and LA size. Moreover, ectopic foci in the initiation of paroxysmal AF are usually found in PVs and, in fact, variant anatomy of PVs is also known to be a predictive factor associated with higher rate of AF recurrence after ablation procedures [44].

Thus, the parameters that we are going to consider in order to evaluate the segmentations are listed below:

- Parameter A: Proper segmentation of all the PVs, including possible abnormalities, such as left common trunk (LCT) or right middle pulmonary vein (RMPV). Here, a good length and width of the veins will be punctuated positively. This aspect is clinically relevant as it gives room to physicians to detect ectopic foci within the PVs.
- Parameter B: Proper segmentation of the LA chamber. As mentioned before, the LA size is an important predictor for AF ablation success.

The punctuation criteria

For Parameter A, all of the PVs of each segmentation were subjectively analyzed taking both diameter and length into consideration. These were the followed steps:

1. Each PV within the LGE-MRIs was delineated and compared to the output segmentation from the HCB U-Net model (see Annex 1):

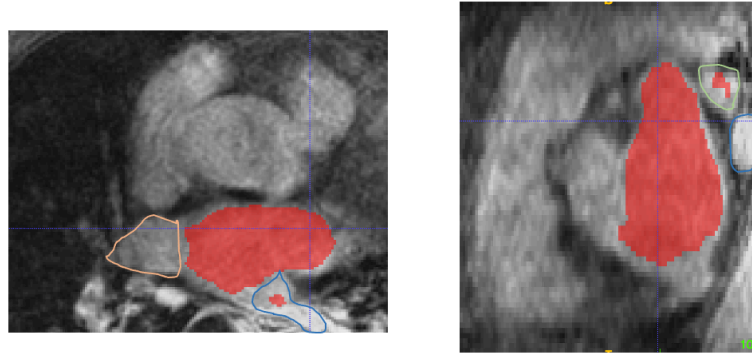


Figure 21. Comparison between the expected segmentation vs. the actual one. In red, the actual predicted segmentation, and in fine lines, the expected one.

2. Then, a subjective punctuation between good (G), average (A), and poor (P) was given for the diameter and length of each of the visible PVs in the MRI.
3. The previous punctuations were converted into numbers ($G = 3$, $A = 2$, $P = 1$), and an average was computed in order to assign an overall punctuation to each inlier segmentation.

For Parameter B, the maximum dimensions between the real LA and the segmented LA seen from the axial view were compared following these steps:

1. Maximum height (from anterior to posterior) and length (from left to right) from the axial plane were computed in each MRI and segmentation, as seen in this example (see Annex 2):

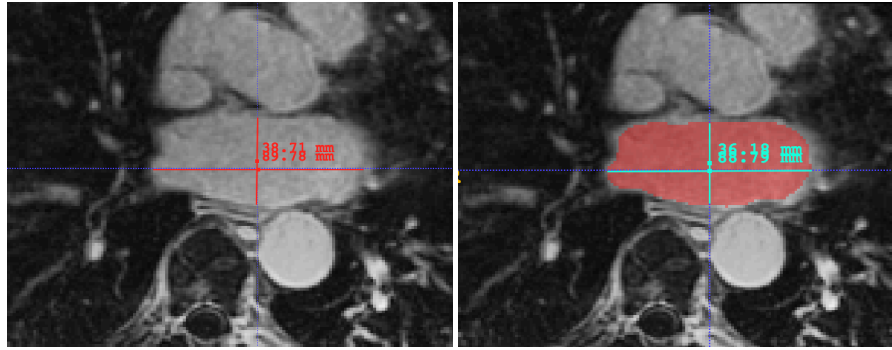


Figure 22. Left: Axial measures of the real LA, Right: Axial measures of the segmented LA.

2. The surface area for each height and length calculation was computed using the ellipse surface formula: $A = \pi \cdot \frac{height}{2} \cdot \frac{length}{2}$ (5)
3. For each patient we computed the difference between the expected and the real LA surface.

5.4.3. Results

Parameter A

Table 2 shows the PVs punctuation for all the inlier data points, i.e. all the patients except for 3, 9, and 19. According to this punctuation system, the best PVs segmentation is the one for Patient 11, whereas the worse is in Patient 6 (Fig. 23).

PATIENT -->	1	2	4	5	6	7	8	10	11	12	13	14	15	16	17	18	20
RSPV																	
diameter		G 3 A 2	G 3 A 2	G 3 P 1	P 1 P 1	G 3 A 2	G 3 A 2	G 3 A 2	G 3 A 2	G 3 A 2	G 3 A 2	G 3 A 2	G 3 A 2	P 1 G 3	G 3 G 3	G 3 G 3	G 3 G 3
length		G 3 A 2	A 2 A 2	P 1 P 1	P 1 P 1	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	P 1 G 3	G 3 G 3	G 3 G 3	G 3 G 3
RIPV																	
diameter		G 3 G 3	G 3 G 3	P 1 A 2	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	P 1 A 2	P 1 A 2	P 1 P 1	P 1 P 1
length		G 3 A 2	A 2 A 2	P 1 G 3	G 3 G 3	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	A 2 A 2	P 1 A 2	P 1 A 2	P 1 P 1	P 1 P 1
LSPV																	
diameter	A 2 A 2			G 3 A 2	A 2 A 2			A 2 G 3	A 2 G 3	A 2 G 3	A 2 G 3	A 2 G 3	A 2 G 3	A 2 G 3	A 2 G 3	G 3 G 3	P 1 P 1
length	G 3 G 3			G 3 P 1	G 3 G 3			G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	G 3 G 3	A 2 G 3	A 2 G 3	G 3 G 3	P 1 P 1
LIPV																	
diameter	G 3 A 2			G 3 A 2	P 1 P 1			A 2 G 3	G 3 G 3		A 2 A 2	A 2 A 2	A 2 A 2	P 1 P 1	P 1 G 3	G 3 G 3	G 3 G 3
length	G 3 A 2			G 3 P 1	P 1 P 1			G 3 G 3	G 3 G 3		G 3 G 3	G 3 G 3	G 3 G 3	A 2 P 1	P 1 G 3	G 3 G 3	G 3 G 3
Abnormalities:																	
RMPV																	
diameter			A 2					G 3							P 1		
length			A 2					A 2							P 1		
RCT																	
diameter	P 1																
length	A 2																
LCT																	
diameter			G 3					G 3		G 3							
length			G 3					G 3		G 3							
Total	14	21	19	22	10	14	17	24	29	10	21	19	22	11	20	20	16
Average	A 2,33	G 2,63	A 2,38	G 2,75	P 1,25	A 1,75	G 2,83	A 2,4	G 2,9	A 1,67	G 2,63	A 2,38	G 2,75	P 1,38	A 2	G 2,5	A 2

Table 2. Punctuation table for each inlier data point. For normal anatomies, the right superior (RSPV), right inferior (RIPV), left superior (LSPV), and left inferior (LIPV) pulmonary veins are punctuated in terms of diameter and length segmentation precision. Also, for abnormal anatomies, the right middle pulmonary vein (RMPV), right common trunk (RCT), and left common trunk (LCT) have been analyzed. Finally, the total and average punctuation have been computed. Empty punctuations may account for inappropriate LA centering crop.

It is important to bear in mind that some degree of inappropriate LA centering crop is also found within the inlier set. This is seen in patients 6, 12, 17, and 20. However, since the mistake does not include more than 10% of the whole volume, these have not been categorized as outliers.

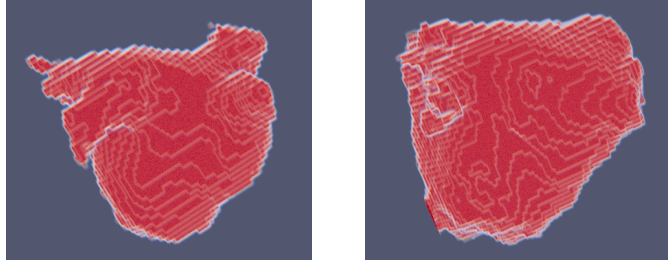


Figure 23. Best (Patient 11, left) and worse (Patient 6, right) PV segmentation according to our punctuation criteria. 3D volume reconstruction seen from a posteroanterior (PA) view.

Further discussion about the obtained data is included in the following section.

Parameter B

Table 3 shows the LA distances and surface areas for all the inlier data points and Patient 9, i.e. all the patients except for 3 and 19. According to this punctuation system, the best LA segmentation is the one for Patient 14, as it differs the less, whereas the worse is in Patient 16, where the difference is the largest.

Patient	Length	Width	Surface area	Segmentation	Length	Width	Surface area	Normalized error
1	99,2	44,09	3435,12	1	90,29	37,99	2694,01	0,22
2	93,55	39,78	2922,80	2	90,96	37,99	2714,00	0,07
4	94,09	41,4	3059,38	4	88,67	39,03	2718,10	0,11
5	112,8	55,82	4945,26	5	107,4	52,06	4391,35	0,11
6	121	45,16	4291,70	6	115,1	41,25	3728,97	0,13
7	91,74	55,93	4029,89	7	86,99	42,82	2925,54	0,27
8	89,78	38,71	2729,56	8	88,34	37,02	2568,52	0,06
9	105,4	61,85	5120,00	9	88,71	45,7	3184,04	0,38
10	106,5	38,71	3237,89	10	103,2	37,64	3050,84	0,06
11	85,54	30,63	2057,81	11	83,23	29,48	1927,07	0,06
12	99,46	47,31	3695,65	12	97,35	45,96	3514,03	0,05
13	90,86	39,25	2800,93	13	86,03	38,71	2615,55	0,07
14	75,27	37,1	2193,24	14	75,25	35,49	2097,50	0,04
15	93,01	45,7	3338,38	15	87,01	44,01	3007,53	0,10
16	121	54,84	5211,62	16	96,78	41,4	3146,85	0,40
17	100,5	45,16	3564,59	17	91,4	41,41	2972,63	0,17
18	89,29	46,7	3274,99	18	80,25	37,83	2384,36	0,27
20	88,71	44,09	3071,87	20	82,64	41,55	2696,82	0,12

Table 3. LA distances (length and width) and LA surface area for all inlier and Patient 9. Last column is the normalized segmentation error with respect to the true LA volume.

It is important to note that the error magnitude is normalized with respect to the LGE-MRI true surface area. Moreover, the units are not included, as only relative measures are considered relevant.

Patient 9 has been considered in this calculation because, even though it is an outlier, it includes the LA blood pool within the segmentation.

5.4.4. Identified problems

Having analyzed the obtained segmentations with the HCB U-Net architecture, these are the main two problems that are predominant in the obtained segmentations:

- PV segmentation inaccuracy: Through our subjective evaluation, it has been seen that only 7 out of the 20 images obtain a Good score on the segmentation of the PVs, 8 obtain an Average score, and 2 obtain a Poor score. Therefore, it is clear that the HCB U-Net is still a little bit far from properly segmenting the PVs in GE Healthcare LGE-MRIs.
- Inappropriate LA centering crop: As mentioned in the outliers section, this problem does not reflect a mistake within the model itself, but rather within the pre-processing of the images. Nonetheless, this problem should be addressed as it clearly has a significant effect on the rest of the pipeline. The images within the inlier data points that are not correctly centered are patients 6, 12, 17 and 20 (Fig. 24).

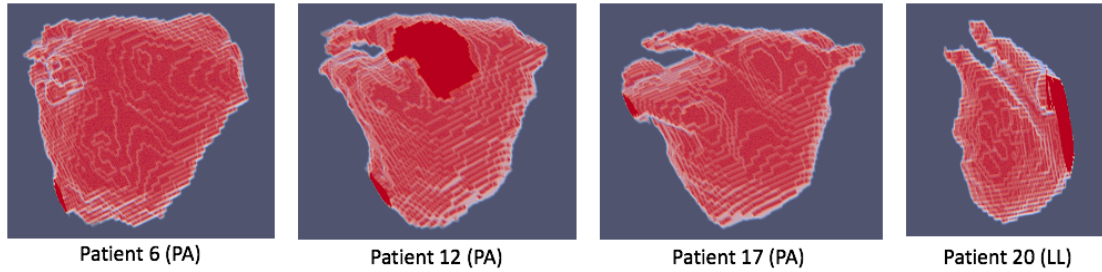


Figure 24. Inappropriately cropped images from which an inappropriately cropped segmentation volume is obtained. The darker red parts are the cropped portions. Posteroanterior views (PA), and Left-Lateral view (LL).

5.4.5. Overcoming the inaccurate PV delineation

Under our perspective, there are two main factors that could cause this poor performance: the use of an inconvenient loss function to train the model, and the fact that the model has learned an average volume threshold that prevents it from fully segmenting the PVs. We will discuss both possible causes.

5.4.5.1. Loss function

The HCB U-Net uses a combination of the Generalized Dice Loss function, and the Boundary Loss function by Kervadec et al. [42] in order to train the model. The PVs have a very narrow anatomy compared to the LA alone, making it difficult for the network to delineate the correct contour of these, that is why, the HCB U-Net uses the Boundary Loss function.

Kervadec's Boundary Loss is known to perform very well in highly unbalanced segmentations. That is, in datasets that have the class imbalance problem, such as the present. This issue, however, has already been tackled from two other points of view:

- The patch based training, that demands the input image to contain a minimum amount of foreground labels.
- The Generalized Dice Loss, which applies a specific weight to each label so that the model does not learn mostly from the background pixels.

That is why, we propose to hypothetically replace Kervadec's boundary loss, with the one proposed by Han et al.'s [42], which is a contour loss that integrates distance, area, and boundary information, as it can be seen in the equation:

$$\mathcal{L}_e = d + \alpha \cdot a + \beta \cdot e \quad (5)$$

Where d , a and e indicate the distance, area and boundary, respectively, and α and β are the weights of the area and the boundary, respectively.

This function works in a way that, if the segmentation boundary of the target is distinct, the weighted constraint of e in the loss function is strong. If segmentation does not need boundary information, the total loss degenerates into Dice loss, that is, losing its utility in optimizing edge segmentation. Since the segmentation boundary of PVs is contrasted with the black background of the images, it will have a strong coefficient within the function.

5.4.5.2. Average volume threshold

As mentioned previously, one of the features that the model may be learning is the amount of volume that it has to segment. This is perfectly possible, especially given the fact that, when patches are formed during the training phase, a percentage of LA volume within the image is established (positive probability percentage). While this is advantageous for the class imbalance issue, it might, at the same time, be teaching the network that all LGE-MRIs contain a similar amount of LA and PVs volume.

This hypothesis is also seen in the collected data during evaluation. On the one hand, it has been observed that, the bigger the true LA volume, the more difficulties the model has to segment it. This can be seen Fig. 25, where there exists a significant positive linear relationship between the true LA volume and the difficulties that the model has to segment it, i.e. the normalized difference, with $p < 0.005$ (within a 95% confidence interval).

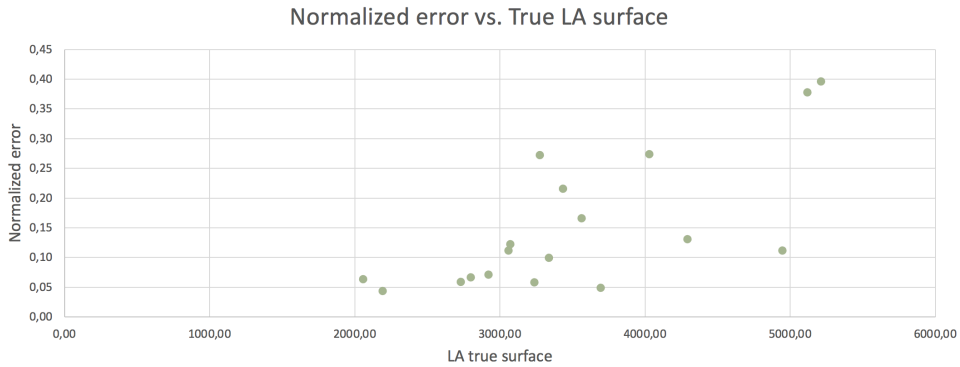


Figure 25. Normalized error vs. True LA surface graph. A positive relationship between the two variables is observed.

On the other hand, we have analyzed the relationship between the size of the true LA, and the PVs subjective punctuation (Parameter A) (Fig. 26). We based this analysis under the hypothesis that maybe, the bigger the true volume of the LA, the harder it is for the model to segment the PVs as, according to our theory, it has invested most of the available segmentation volume in segmenting the LA. Even though a negative linear relationship is glimpsed through the graph, it is not considered as statistically significant as $p > 0.05$.

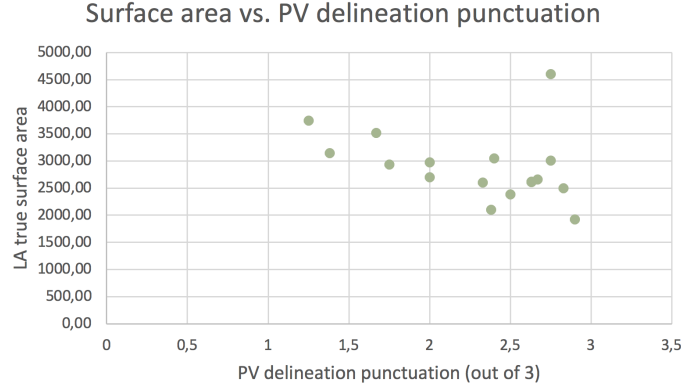


Figure 26. True LA surface area vs. PV delineation subjective punctuation. A slight negative relationship between the two variables is observed.

To overcome this problem, we consider that feeding the model with a more variable dataset would help overcome the present issue. This could be done by dynamically changing the positive probability percentage (PPP) during patch extraction, in a way that the network sees a wider range of LA volumes. By dynamic we mean that, for each epoch, an heterogeneous distribution of PPPs across the input dataset is established.

In fact, J.Reventós experimented the performance of the model for different positive probability percentage magnitudes, but always in a uniform way throughout the whole dataset, and never using a combination of it.

5.4.5.3. Lack of context learning

Another possible cause that would explain the inappropriate PV segmentation is the lack of the model's context learning. It has been observed throughout our output dataset, that the model properly learns to segment the LA, but it almost never prioritizes the PVs. That is to say, that the model has learned to segment up to the average volume threshold, but it has not learned the importance of properly segmenting the PVs. That could probably account for a lack of context learning, as the HCB U-Net is not fully understanding the context.

Context learning consists in having the ability to extract proper feature hierarchy. To achieve this within our network, we propose a solution inspired in the Bian et al. [42] pyramid pooling module (Fig. 27). This module is capable of extracting information from multiple scales and enlarge receptive fields, ultimately promoting the segmentation in fine scales, as it happens in the PVs. The module consists in 4 pooling layers that aggregate context information from global to fine scales.

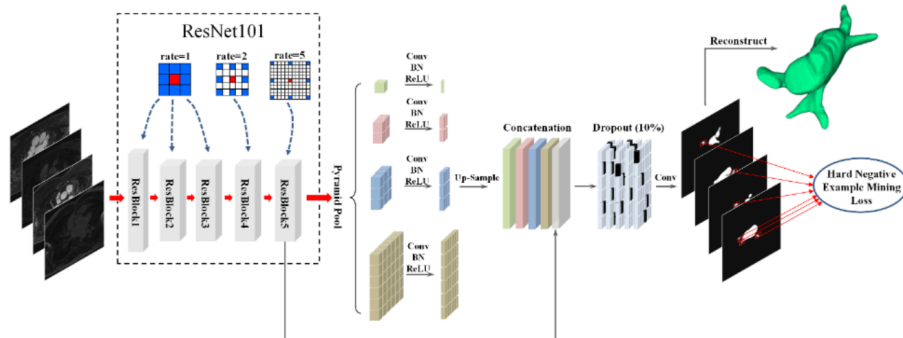


Figure 27. Schematic view of the pyramid pooling module (dotted square) and the framework where it is embedded (Bian et al.).

Bian et al. use this solution for the LA segmentation problem, but outside of a U-Net framework. Therefore, the compatibility of the module with a U-Net scheme should be tested in case we wanted to adopt it for the HCB U-Net model.

5.4.6. Overcoming the inappropriate LA centering crop

The inappropriate LA centering crop does not require much analysis in order to be detected. It is very easily observed and the prevalence is high. Out of the 20 studied 3D LGE-MRIs, 5 of them are not correctly centered: subjects 6, 9 (outlier), 12, 17 and 20.

The fact that 25% of the images are not well cropped represents a critical problem. Here, we discuss possible solutions that could help overcome this issue.

5.4.6.1. *Multi-CNN for dynamic LA centering crop*

It has been presented in previous sections the use of a Multi-CNN scheme, in which there are two consecutive networks (as opposed to one): the first one localizes the LA on each image to exclusively crop background pixels, and the second one focuses on the segmentation task *per se*.

This approach would clearly improve the cropping task, as it would dynamically change the center according to the anatomy of each patient. The downside to this approach is the high computational cost of running two networks.

5.4.6.2. *Increasing the security margin*

Increasing the security margin means that, from the current established LA center, the crop would output a bigger image than the current one. In other words, bigger images would be inputted into the network, and therefore, more memory and time would also be required. Nevertheless, this time and memory increase would maybe be smaller than the one of running a Multi-CNN.

5.4.6.3. *Localizing the descending aorta*

The descending aorta (DA) represents a clear anatomical structure from the axial view (Fig. 28), as it is an almost perfectly round circle with a distinctive intensity value. Moreover, it has a pretty consistent shape throughout the patients' spectrum and always located very close to the right lower corner of the desired cropped image.

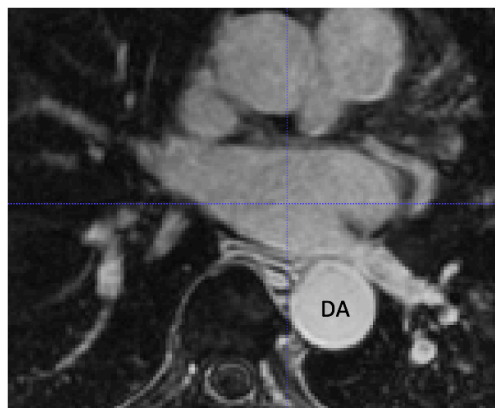


Figure 28. Axial view of an LGE-MRI, where the DA is clearly identified.

On the other hand, transfer learning is the discipline within machine learning where a model developed for a task is reused as the starting point for a model on a second task.

Therefore, a pre-trained circle detection model could be used as the starting point of the segmentation network. Then, a simple algorithm would have to use this information to crop the image into a fixed size, and finally the reduced image would be inputted into the HCB U-Net.

The advantage of this method is the fact that only a small amount of time and computer memory would have to be invested into training the pre-trained model. The challenge, however, would be that there are other circular shapes in the axial plane, though not as perfectly circular.

This is a completely novel strategy that has not been tested in any LA segmentation algorithm, and would be interesting to test.

6. EXECUTION CRONOGRAM AND FINANCIAL VIABILITY

In this section we basically define the tasks, timing, and cost of this project. First, the tasks have been broken down hierarchically using a Work Breakdown Structure (WBS) scheme and explained in detail. Then, a precedence analysis and a GANTT diagram are presented, so that the reader gets a clear perspective of the work and timings within the project. Finally, based on the task breakdown, we make an analysis of the cost of the project.

6.1. WORK BREAKDOWN STRUCTURE

A hierarchical decomposition of the total scope of the work is presented below:

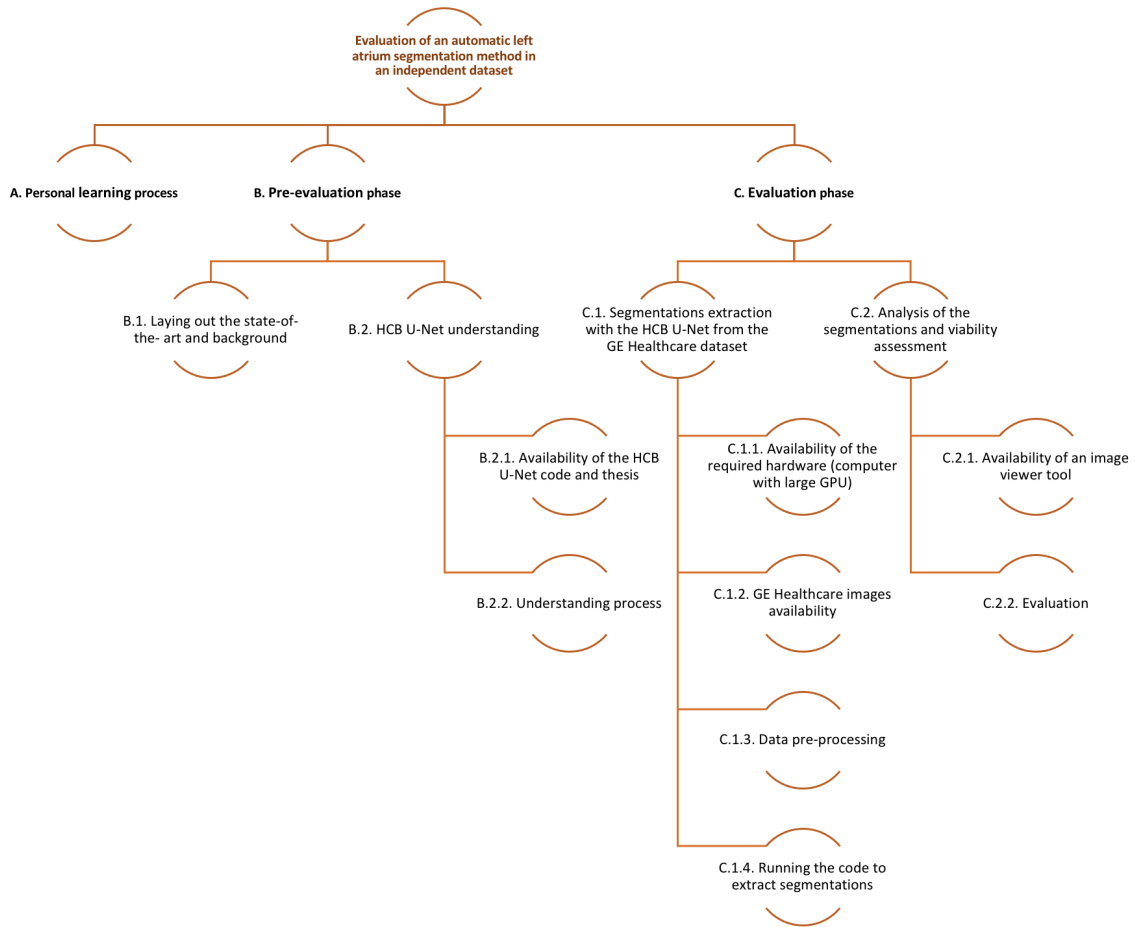


Figure 29. Work Breakdown Structure of the present project.

Evaluation of an automatic left atrium segmentation method in an independent dataset is the ultimate objective of this project. From here, we will explain each sub-task in slightly more detail.

- A. Personal learning process:** Before undertaking the present project, I personally barely knew anything about artificial intelligence and its workings. To acquire some knowledge regarding that field, I took two virtual courses on the subject via Coursera:
- Neural Networks and Deep Learning [48] (From December 1st to 31st 2020)
 - AI for Medical Diagnosis [49] (From February 1st to 28th 2021)

- B. Pre-evaluation phase:** Each neural network has its own ins and outs, and getting a good grip of them is complicated. To understand the HCB U-Net model, I had to delve into a lot of literature, make sense of the code, and read J.Reventós's Final Master Thesis.
 - B.1. Laying out the state-of-the-art and background:** Research about existing similar models, DL advancements, LA anatomy, and other subjects has been a constant over the course of this thesis. (From 1st October 2020 to 1st June 2021)
 - B.2. HCB U-Net understanding:** To have a good understanding of the HCB U-Net model, we had to thoroughly understand the code as well as the written thesis. Thus, this step can be divided into these two parts:
 - B.2.1. Availability of the HCB U-Net code and thesis (From April 15th 2021)
 - B.2.2. Understanding process (From April 15th to May 1st 2021)
- C. Evaluation phase:** This is the purely evaluation phase, and it consists of several steps:
 - C.1. Segmentations obtention with the HCB U-Net from the GE Healthcare dataset:** This step was carried out by Gaspar Delso, one of this thesis supervisors. These are the requirements and steps followed for this phase:
 - C.1.1. Availability of the required hardware (computer with large GPU): The model was trained on a NVIDIA GeForce RTX 2070 GPU with 8GB VRAM. For testing, similar capabilities were required. (From February 1st 2021)
 - C.1.2. GE Healthcare images availability. (From February 1st 2021)
 - C.1.3. Data pre-processing: The images were cropped before being fed to the model. (May 12th 2021)
 - C.1.4. Running the code to obtain segmentations (May 12th 2021)
 - C.2. Analysis of the segmentations and viability assessment:** This step had fewer requirements than the previous one as the evaluation was done from a qualitative approach.
 - C.2.1. Availability of an image viewer tool (ITK Snap) (From November 1st 2020)
 - C.2.2. Evaluation: this is the main analytical step of the project (From May 12th to June 9th 2021)

6.2. PRECEDENCE ANALYSIS

In this section we present a table (Table 4) that shows the dependencies between the tasks mentioned above. Since they have been broken down in a hierarchical way, we will only show dependencies between lower level tasks.

ACTIVITY	NAME	ANTECEDENT	DURATION
A	Personal learning process	-	3 months
B.1	State-of-the-art and background	-	8 months
B.2.1	Availability of the HCB U-Net code and thesis	-	1 day
B.2.2	Understanding process	A / B.2.1	15 days
C.1.1	Availability of the required hardware	-	1 day
C.1.2	GE Healthcare images availability	-	1 day
C.1.3	Data-preprocessing	C.1.1 / C.1.2	1 day
C.1.4	Obtaining segmentations	B.2.1 / C.1.3	1 day
C.2.1	Availability of an image viewer tool	-	1 day
C.2.2	Evaluation	B.1 / B.2.2 / C.1.4 / C.2.1 /	1 month

Table 4. Precedence analysis of the present project.

6.3. GANTT CHART

In this section we illustrate the project schedule through a Gantt chart and then specify the cost of each of each task.

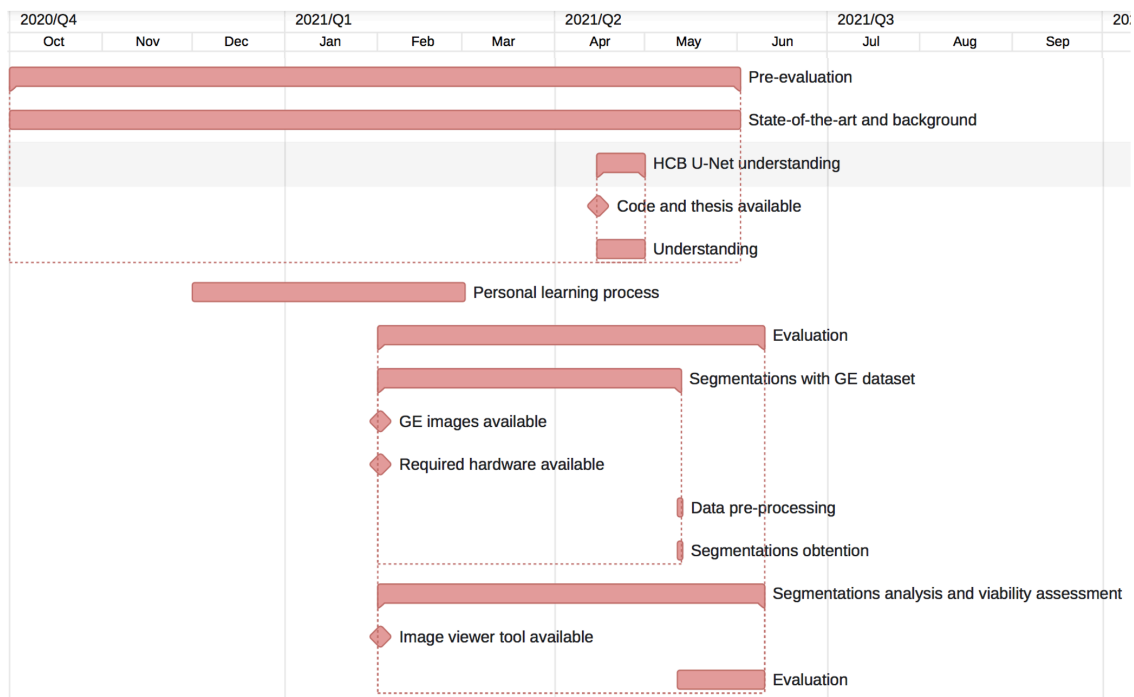


Figure 30. Gantt chart of the project. The rhombus shape represents the point in time when a specific resource becomes available.

6.4. FINANCIAL VIABILITY

To compute the cost of the project in order to assess its financial viability, we are going to adhere to some assumptions:

- Costs will be computed supposing that we are a company outside of the hospital that hires a research group within the hospital to do this research project, so that the company can use this information to create their product later on. The deal is a “win-

win” as the company gets patient anonymized data from the hospital, and the hospital will have an advantageous access to the product once it reaches the market.

- The personnel cost will be reflected as an 8€/hour hourly fee.
- The HCB U-Net code and written thesis are sold to our company for 2,000€.

2

Phase	Duration	Hours/Period	Total Hours	Total Cost	Comments
A Basic learning				400,00 €	
Tuition fee				80,00 €	
Learning	8 weeks	5 hours/week	40 hours	320,00 €	
B Pre-evaluation				3.200,00 €	
Background research	34 weeks	4 hours/week	136 hours	1.088,00 €	
U-Net code and thesis				2.000,00 €	
Understanding the HCB U-Net	2 weeks	7 hours/week	14 hours	112,00 €	
C Evaluation				947,00 €	
NVIDIA GeForce RTX 2070 GPU with 8GB VRAM				499,00 €	
GE images				0,00 €	Granted by the hospital
Pre-processing and segmentations	1 day	8 hours/day	8 hours	64,00 €	
ITK Snap				0,00 €	Free license
Evaluation	4 weeks	12 hours/week	48 hours	384,00 €	
TOTAL			246 hours	4.547,00 €	

Figure 31. Breakdown of costs of the project.

All in all, the whole project is pretty affordable as it is usually the case for software solutions, given that they do not require any hardware expenditure. It can also be observed that most of the cost comes from research (we include the HCB U-Net code and thesis), which is also what usually happens in medical software devices.

² Hardware cost [58]

7. TECHNICAL VIABILITY

This chapter presents a strategic analysis of the project aiming to assess the technical viability of it. We will do this by means of the SWOT matrix technique, in which Strengths, Weaknesses, Opportunities, and Threats are independently analyzed.



Figure 32. SWOT analysis of this project.

7.1. STRENGTHS

- **Less time consuming than current methods:** Current segmentation methods require a big amount of human interaction, preventing healthcare professionals (physicians, engineers, etc) from spending time doing more specialized tasks. Therefore, implementing an automatic LA segmentation tool in, for example, the Arrhythmias Section at HCB, would represent a significant upgrading to the workers.
- **Less prone to errors than current methods:** The performance of manually segmenting the LA from LGE-MRIs largely depends on the experience of the professional that does the task. Replacing this method with an automatic tool, should significantly reduce the probability of making errors during the process.
- **Cost-effectiveness:** The new automatic segmentation model is extremely cost-effective, as it would considerably reduce workforce and time, at the same time that performance would be improved.
- **Versatility:** It has been demonstrated in this thesis that the HCB U-Net model is very versatile in terms of the type of data that it can process.

7.2. WEAKNESSES

- **Black box effect:** There is always the same predicament when it comes to implementing AI schemes, which is the black box effect. It means that AI systems are impenetrable even by the designer as the process is largely self-directed by the machine. The importance of this effect is even higher in the medical field, where the decision made by the machine can have life-threatening consequences. Therefore, it is reasonable to think that healthcare professionals that are used to transparent methods, could be reluctant to such an opaque tool.
- **Inter-subject variability:** As aforementioned, manually segmenting the LA requires a lot of expertise, and not all the professionals agree on which structures should be delineated

and which others should be left out. This is important when training a DL model as it bases its knowledge on the provided ground truth. Therefore, the performance measures of the HCB U-Net may not be 100% objective.

7.3. OPPORTUNITIES

- **Relatively easy funding:** Not only AF is the most common type of arrhythmia worldwide, but also cardiovascular diseases represent a huge problem in developed countries nowadays. Moreover, catheter ablation procedures are becoming a very popular form of treatment for AF, and more and more professionals seek for an automatic tool for the segmentation of the LA. Thus, it is safe to say that getting funding for this project should not be extremely complicated.
- **Desirable environment:** HCB and especially the Arrhythmias Section perfectly qualify for a research and innovation friendly environment. The department has a very powerful combination of biomedical engineers, physicians, and the presence of companies, all of them involved in research projects and eager to innovate. This is a clear competitive advantage for this project.

7.4. THREATS

- **Abundant competition:** The automation of the LA segmentation is clearly a fashionable topic. That is why a large number of research groups are putting their efforts into finding the appropriate model. The high competitiveness in the field clearly poses a threat to our project, especially since our model does not provide fibrotic tissue segmentation. The rest of possible threats are equal for the rest of the competition.

8. REGULATIONS AND LEGAL ASPECTS

If we hypothetically were to launch the modified HCB U-Net to the European market, some important regulatory and legal aspects would have to be taken into account. In fact, up until very recently, the United States Food and Drug Administration (FDA) posed a more strict regulation for medical devices than the European Union (EU). However, as of May 26, 2021, the EU Medical Device Regulation (EU MDR) replaced the former EU Medical Device Directive, imposing stringent regulatory requirements that need to be met before any medical device (including AI software tools) can be used in clinical practice [50].

In this chapter, we explain the steps that would have to be followed if we wanted to convert the modified HCB U-Net into a market product within the EU, presenting at the same time the existing regulations and required standards for such a device [51].

The EU MDR applies to processing software that provides a diagnosis or therapy by itself, but also to software that merely gives information intended to inform a medical professional in making the final diagnostic or therapeutic decision. In this example, our software device aims to guide the medical professional instead of directly providing a determining decision.

Within the EU, the EU MDR legislation is the most stringent and demanding one regarding medical device software (MDSW), but as developers, we should also be aware of other EU laws and guidance documents supporting their implementation. For that matter, we should pay special attention to the General Data Protection Regulation (GDPR) [52], as our device would collect personal data. Listed below, are the main documents that we should keep in mind:

- Regulation (EU) 2017/745: MDR: Medical Device Regulation (MDR)
- MDCG 2019-11: Guidance on Qualification and Classification of Medical Device Software (MDSW)
- MDCG 2019-16: Guidance on cybersecurity for medical devices.
- MDCG 2020-1: Guidance on clinical evaluation of MDSW.
- Regulation (EU) 2016/679: General Data Protection Regulation (GDPR) for the protection of natural persons with regard to the processing of personal data and on the free movement of such data.

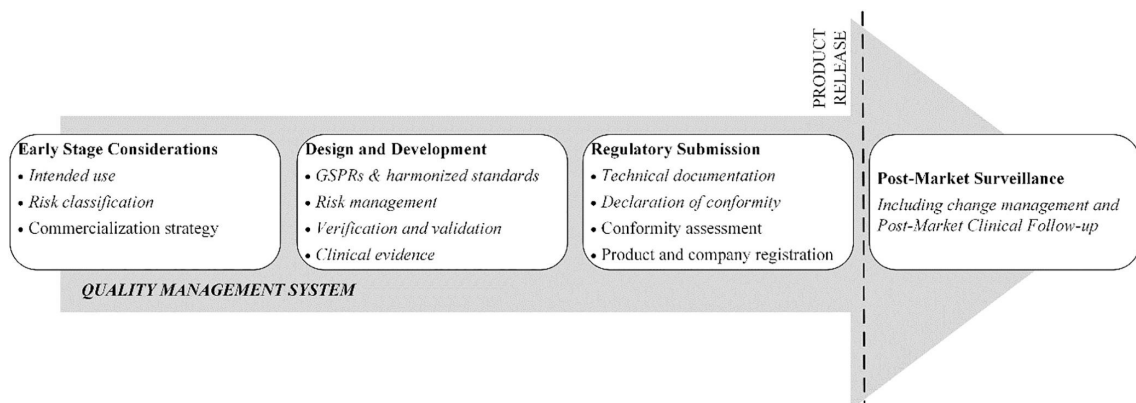


Figure 33. Overview of the regulatory roadmap for commercial and in-house MDSW. Italics represent the steps that the two products have in common. Non-italics is just for commercial use.

The EU MDR regulates both commercial and in-house use intended devices, being more stringent on the first type of products. An overview of the regulatory roadmap for these two types of MDSW is shown in Fig. 33.

8.1. EARLY STAGE AND CONSIDERATIONS

In this stage, as manufacturers, we would have to explain the intended use of our MD, and classify it within the EU MDR classification scheme.

This classification method consists of several rules. While MDSW do not directly relate to many of these rules, as they refer to physical aspects of the device, they do relate to the consequence of indirect harm from failure to provide correct information. Therefore, a new rule (Rule 11) has been introduced into the MDR in order to address the risks related to the information provided by MDSW.

Rule 11 categorizes patient management, which is the type of information provided by the device to the clinical decision, and the patient condition, i.e. the clinical situation. More detailed, Rule 11 states [53]:

Software intended to provide information which is used to take decisions with diagnosis or therapeutic purposes is classified as class IIa, except if such decisions have an impact that may cause:

- *death or an irreversible deterioration of a person's state of health, in which case it is in class III; or*
- *a serious deterioration of a person's state of health or a surgical intervention, in which case it is classified as class IIb.*

Software intended to monitor physiological processes is classified as class IIa, except if it is intended for monitoring of vital physiological parameters, where the nature of variations of those parameters is such that it could result in immediate danger to the patient, in which case it is classified as class IIb.

All other software is classified as class I.

Finally, for this stage, a commercialization strategy has to be presented in case it wants to enter the market.

8.2. DESIGN AND DEVELOPMENT

During the design and development phase, the first thing to do is identify with which General Safety and Performance Requirements (GSPR) our product will have to comply. For MDSW like ours, we should pay special attention to the following standards:

- IEC 82304-1:2016, it applies to the safety and security of health software products designed to operate on general computing platforms and intended to be placed on the market without dedicated hardware, and its primary focus is on the requirements for manufacturers. It covers the entire lifecycle including design, development, validation, installation, maintenance, and disposal of health software products [54].
- IEC/EN 62304 Medical Device - Software Life Cycle Processes: it defines requirements for the life cycle of the development of medical software and for software within medical devices [55].

Another core activity within the design and development phase is risk management. The ISO 14971 [56] standard explains the required application of risk management to medical devices

Finally, regardless of their application, all medical devices require clinical evidence, which can be obtained in a wide range of ways.

8.3. REGULATORY SUBMISSION

The last step before releasing the product and obtaining the CE mark is the regulatory submission step, which consists in delivering the technical documentation and writing the Declaration of Conformity, where the manufacturer confirms that the product complies with the EU MDR.

Moreover, if we want the product to reach the European market, the device and the manufacturer's quality management system have to meet the requirements of the regulation so that the conformity assessment can be completed. Also, the product and company have to be registered.

8.4. POST-MARKET SURVEILLANCE

Once the CE certificate has been issued, post-market surveillance aims to continuously verify the benefits of a medical device and to identify previously unknown risks by observing daily usage.

9. CONCLUSIONS

The objective of the project was to evaluate the performance of the automatic LA segmentation model HCB U-Net when tested with an independent set of images. This was accomplished by qualitatively evaluating the predicted segmentations with respect to my personal knowledge about the anatomy of the left atrium. The performance assessment of the model on the twenty 3D LGE-MRI images from the GE dataset put forward some interesting discussion points.

In the first place, it should be emphasized that three out of the twenty analyzed segmentations were considered outliers, as they did not comply with the acceptance criterion. One of them showed inappropriate left atrial centering crop caused in the pre-processing stage, and only about 50% of the segmentation could be appreciated. The second outlier data point included an extra anatomical structure, in this case, the pulmonary artery. Also in that outlier, the pulmonary veins were completely ignored within the segmentation. The last outlier was carried out in a low quality image and the segmented portion was completely incoherent.

For the inlier data points, two clinically relevant parameters were used to assess the quality of the segmentations. On the one hand, we measured how well the pulmonary veins were segmented by setting a qualitative scale; and, on the other, we measured the surface of the left atrium on the axial plane and compared it to the segmented portion of that chamber.

With this data and after closely analyzing each segmentation, we observed two persistent problems in them: the inaccurate pulmonary veins delineation, and the inappropriate image crop.

Regarding the first issue, the inaccurate pulmonary veins delineation, we found a statistically significant relationship between the true left atrial volume size, and the difficulties that the model had to segment it. Hence, the bigger the true left atrium, the harder it was for the network to include the whole portion within the segmented mask. We suspect that this is because the model sets a threshold volume based on the images it has learned from, in which it ends the segmentation. This is a very critical problem because the patients that have a bigger left atrial volume, are also the ones with a higher risk of AF recurrence after a catheter ablation procedure. To solve this, we propose to change the loss function, and to increase the variability of the input data during the training phase of the HCB U-Net model, particularly by dynamically changing the positive probability percentage of the patches.

As for the inappropriate image crop issue, we observed that 25% of the tested images were not fully well cropped during the data pre-processing phase. This problem was very easily observed, and three strategies were proposed to overcome it. First, we suggested using a dynamic LA centering crop architecture by means of a Multi-CNN structure. Also, we thought of increasing the security margin of the cropped images, so that the probability of including the whole LA and PVs aggregate increases. Finally, we proposed a transfer learning strategy for the detection of the descending aorta, which has a very distinguishable shape in the axial plane. Overall, the first and second solutions would be a safe option, as they have shown good results in previous experiments. The downside to them, however, is the need for more computational capacity, which is not always easy to access. The third option has never been

tested and it is an interesting technique to keep in mind even for other uses in the medical image segmentation field.

All in all, it could be said that the HCB U-Net is a promising tool, though still lacking a little bit of adaptability when it comes to independent datasets, especially in the LA centering crop feature. Moreover, the catheter ablation procedure for atrial fibrillation patients is a very widespread technique nowadays, and a lot of money is being invested in this field, making it pretty viable for our solution to reach the market. For that matter, if the model were to become a market product, we think it should do it as a combination of an automatic and manual tool, so that once the automatic model has output the segmentation, the user can manually edit and modify it as he/she considers appropriate.

Very importantly, we consider that the main future working line of this project should consist in adding a module to the network that is capable of detecting scar tissue within the left atrium and the pulmonary veins.

It is worth mentioning also, that some limitations and roadblocks have been found during the execution of this project. On one side, it must be said that, even though a lot of work has been put into evaluating the segmented anatomies, my personal knowledge about the anatomy of the left atrium is minimal with respect to the knowledge from experienced professionals in the field. That may have limited the accuracy of results obtained in this work. Moreover, in the absence of ground truth segmentations, we have not been able to obtain comparable metrics from the predicted segmentations. Nonetheless, since there were many clinically valuable qualitative parameters to analyze, a lot of meaningful information has been successfully extracted from the data. Last but not least, a general burden amongst research projects nowadays is the Covid-19 pandemic. In my case, the fact of not being able to physically witness how the Arrhythmias Section in Hospital Clínic de Barcelona works has been limiting, though not a determining factor on the project's outcomes.

Finally, I want to conclude by stating that this project has been extremely fulfilling for my personal growth in the image processing and artificial intelligence fields, as well as in my ability to carry out complex projects.

- [1] M. Zoni-Berisso, F. Lercari, T. Carazza, and S. Domenicucci, "Epidemiology of atrial fibrillation: European perspective," *Clin. Epidemiol.*, vol. 6, no. 1, pp. 213–220, Jun. 2014.
- [2] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully Automatic Left Atrium Segmentation From Late Gadolinium Enhanced Magnetic Resonance Imaging Using a Dual Fully Convolutional Neural Network," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 515–524, Feb. 2019.
- [3] C. McGann *et al.*, "Atrial fibrillation ablation outcome is predicted by left atrial remodeling on MRI," *Circ. Arrhythmia Electrophysiol.*, vol. 7, no. 1, pp. 23–30, 2014.
- [4] A. Njoku *et al.*, "Left atrial volume predicts atrial fibrillation recurrence after radiofrequency ablation: A meta-analysis," *Europace*, vol. 20, no. 1, pp. 33–42, Jan. 2018.
- [5] G. Lippi, F. Sanchis-Gomar, and G. Cervellin, "Global epidemiology of atrial fibrillation: An increasing epidemic and public health challenge," *Int. J. Stroke*, vol. 16, no. 2, pp. 217–221, Feb. 2021.
- [6] M. H. Kim, S. S. Johnston, B. C. Chu, M. R. Dalal, and K. L. Schulman, "Estimation of total incremental health care costs in patients with atrial fibrillation in the united states," *Circ. Cardiovasc. Qual. Outcomes*, vol. 4, no. 3, pp. 313–320, May 2011.
- [7] J. S. Alpert *et al.*, "2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society," 2014.
- [8] D. Corradi, "Atrial fibrillation from the pathologist's perspective," *Cardiovascular Pathology*, vol. 23, no. 2. Elsevier, pp. 71–84, 01-Mar-2014.
- [9] T. H. Hauser, D. C. Peters, J. V. Wylie, and W. J. Manning, "Evaluating the left atrium by magnetic resonance imaging.," *Europace : European pacing, arrhythmias, and cardiac electrophysiology : journal of the working groups on cardiac pacing, arrhythmias, and cardiac cellular electrophysiology of the European Society of Cardiology*, vol. 10 Suppl 3, no. suppl_3. Oxford Academic, pp. iii22–iii27, 01-Nov-2008.
- [10] A. Doltra, B. Amundsen, R. Gebker, E. Fleck, and S. Kelle, "Emerging Concepts for Myocardial Late Gadolinium Enhancement MRI," *Curr. Cardiol. Rev.*, vol. 9, no. 3, pp. 185–190, Aug. 2013.
- [11] A. D. Dongare, R. R. Kharde, and A. D. Kachare, "Introduction to Artificial Neural Network," 2008.
- [12] C. T. Metz, M. Schaap, A. C. Weustink, N. R. Mollet, T. Van Walsum, and W. J. Niessen, "Coronary centerline extraction from CT coronary angiography images using a minimum cost path approach," *Med. Phys.*, vol. 36, no. 12, pp. 5568–5579, 2009.
- [13] C. Feng, S. Zhang, D. Zhao, and C. Li, "Simultaneous extraction of endocardial and epicardial contours of the left ventricle by distance regularized level sets," *Med. Phys.*, vol. 43, no. 6, pp. 2741–2755, 2016.
- [14] J. C. Bezdek, L. O. Hall, and L. P. Clarke, "Review of MR image segmentation techniques using pattern recognition," *Med. Phys.*, vol. 20, no. 4, pp. 1033–1048, 1993.
- [15] M. R. Kaus, J. von Berg, J. Weese, W. Niessen, and V. Pekar, "Automated segmentation of the left ventricle in cardiac MRI," *Med. Image Anal.*, vol. 8, no. 3, pp. 245–254, Sep. 2004.
- [16] H. C. Van Assen *et al.*, "Cardiac LV segmentation using a 3D active shape model driven

- by fuzzy inference," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2878, pp. 533–540, Nov. 2003.
- [17] I. Išgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. Van Ginneken, "Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in CT scans," *IEEE Trans. Med. Imaging*, vol. 28, no. 7, pp. 1000–1010, Jul. 2009.
 - [18] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Nov. 2014.
 - [19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol. 9351, pp. 234–241.
 - [20] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, 2016, pp. 565–571.
 - [21] "(14) (PDF) From Research to Market: what the EU can learn from the USA." [Online]. Available: https://www.researchgate.net/publication/333759358_From_Research_to_Market_what_the_EU_can_learn_from_the_USA. [Accessed: 13-Jun-2021].
 - [22] J. Kornej, C. S. Börschel, E. J. Benjamin, and R. B. Schnabel, "Epidemiology of Atrial Fibrillation in the 21st Century Novel Methods and New Insights," *Circ. Res.*, vol. 127, pp. 4–20, 2020.
 - [23] S. Ganatra *et al.*, "Efficacy and safety of catheter ablation for atrial fibrillation in patients with cancer," *Eur. Heart J.*, vol. 41, no. Supplement_2, Nov. 2020.
 - [24] "Global Atrial Fibrillation Devices Industry." [Online]. Available: <https://www.reportlinker.com/p05961346/Global-Atrial-Fibrillation-Devices-Industry.html>. [Accessed: 13-Jun-2021].
 - [25] "ADAS 3D: Fibrosis Imaging for the EP Lab." [Online]. Available: <https://www.adas3d.com/en/index.html>. [Accessed: 13-Jun-2021].
 - [26] "2018 Atrial Segmentation Challenge – Atrial Segmentation Challenge." [Online]. Available: <https://atriaseg2018.cardiacatlas.org/>. [Accessed: 13-Jun-2021].
 - [27] "Galgo Medical: Predictive Software for Safer Treatment." [Online]. Available: <https://www.galgomedical.com/en/index.html>. [Accessed: 13-Jun-2021].
 - [28] "Cardiac MRI and CT Software – Circle Cardiovascular Imaging - Home." [Online]. Available: <https://www.circlecvi.com/>. [Accessed: 13-Jun-2021].
 - [29] "Medical Imaging Cloud AI - Arterys." [Online]. Available: <https://arterys.com/>. [Accessed: 13-Jun-2021].
 - [30] "EnSite Precision Cardiac Mapping System Product Overview | Abbott." [Online]. Available: <https://www.cardiovascular.abbott/us/en/hcp/products/electrophysiology/mapping-systems/ensite.html>. [Accessed: 13-Jun-2021].
 - [31] K. Jamart, Z. Xiong, G. D. Maso Talou, M. K. Stiles, and J. Zhao, "Mini Review: Deep Learning for Atrial Segmentation From Late Gadolinium-Enhanced MRIs," *Front.*

- [32] Q. Xia, Y. Yao, Z. Hu, and A. Hao, "Automatic 3D Atrial Segmentation from GE-MRIs Using Volumetric Fully Convolutional Networks," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11395 LNCS, pp. 211–220.
- [33] S. Vesal, N. Ravikumar, and A. Maier, "Dilated Convolutions in Neural Networks for Left Atrial Segmentation in 3D Gadolinium Enhanced-MRI," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11395 LNCS, pp. 319–328.
- [34] C. Bian *et al.*, "Pyramid Network with Online Hard Example Mining for Accurate Left Atrium Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11395 LNCS, pp. 237–245, Dec. 2018.
- [35] C. Li *et al.*, "Attention Based Hierarchical Aggregation Network for 3D Left Atrial Segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11395 LNCS, pp. 255–264.
- [36] X. Yang *et al.*, "Combating Uncertainty with Novel Losses for Automatic Left Atrium Segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11395 LNCS, pp. 246–254, Dec. 2018.
- [37] C. Chen, W. Bai, and D. Rueckert, "Multi-Task Learning for Left Atrial Segmentation on GE-MRI," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11395 LNCS, pp. 292–301, Oct. 2018.
- [38] "Cardiac Atlas Project." [Online]. Available: <https://www.cardiacatlas.org/>. [Accessed: 13-Jun-2021].
- [39] D. Borra, A. Andalò, M. Paci, C. Fabbri, and C. Corsi, "A fully automated left atrium segmentation approach from late gadolinium enhanced magnetic resonance imaging based on a convolutional neural network," *Quant. Imaging Med. Surg.*, vol. 10, no. 10, pp. 1894–1907, Jul. 2020.
- [40] É. Puybureau *et al.*, "Left Atrial Segmentation in a Few Seconds Using Fully Convolutional Network and Transfer Learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11395 LNCS, pp. 339–347.
- [41] Y. Han, X. Li, B. Wang, and L. Wang, "Boundary loss-based 2.5D fully convolutional neural networks approach for segmentation: A case study of the liver and tumor on computed tomography," *Algorithms*, vol. 14, no. 5, p. 144, May 2021.
- [42] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. Ben Ayed, "Boundary loss for highly unbalanced segmentation," in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning*, 2019, vol. 102, pp. 285–296.
- [43] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10553 LNCS, pp. 240–248, Jul. 2017.
- [44] S. Istratoaie *et al.*, "The Impact of Pulmonary Vein Anatomy on the Outcomes of Catheter Ablation for Atrial Fibrillation."

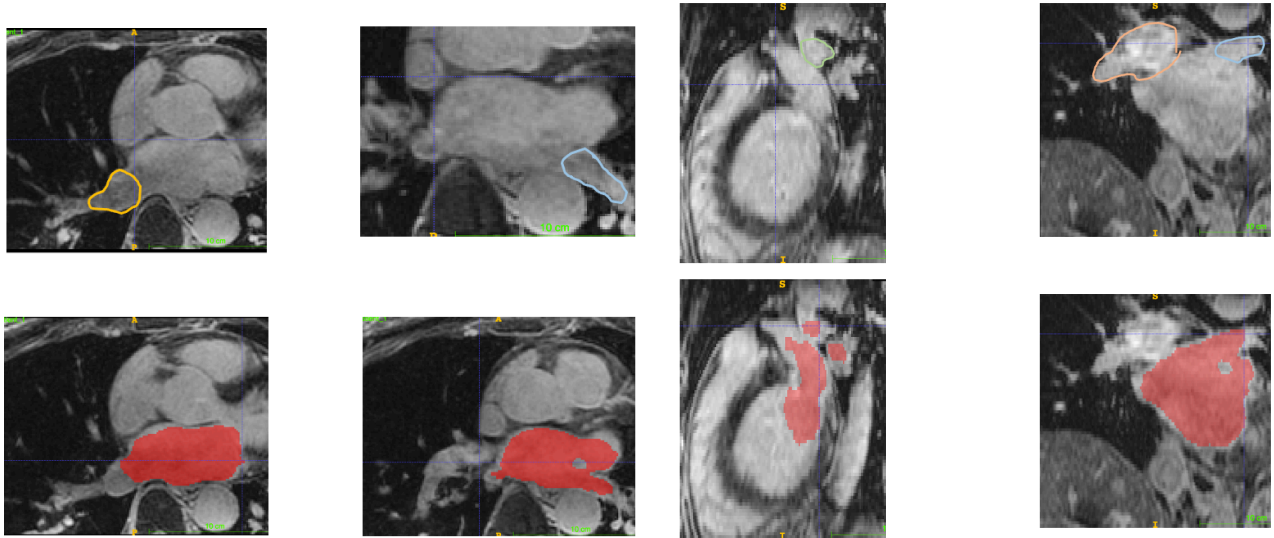
- [45] M. Linhart *et al.*, “Delayed Gadolinium Enhancement Magnetic Resonance Imaging Detected Anatomic Gap Length in Wide Circumferential Pulmonary Vein Ablation Lesions Is Associated With Recurrence of Atrial Fibrillation,” *Circ. Arrhythm. Electrophysiol.*, vol. 11, no. 12, p. e006659, Dec. 2018.
- [46] P. A. Yushkevich *et al.*, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.
- [47] “ParaView: An End-User Tool for Large Data Visualization.” [Online]. Available: https://www.researchgate.net/publication/247111133_ParaView_An_End-User_Tool_for_Large_Data_Visualization. [Accessed: 13-Jun-2021].
- [48] “Redes neurales y aprendizaje profundo | Coursera.” [Online]. Available: <https://www.coursera.org/learn/neural-networks-deep-learning>. [Accessed: 13-Jun-2021].
- [49] “AI for Medical Diagnosis | Coursera.” [Online]. Available: <https://www.coursera.org/learn/ai-for-medical-diagnosis>. [Accessed: 13-Jun-2021].
- [50] “EU MDR – Regulation (EU) 2017/745.” [Online]. Available: <https://eumdr.com/>. [Accessed: 13-Jun-2021].
- [51] R. Beckers, Z. Kwade, and F. Zanca, “The EU medical device regulation: Implications for artificial intelligence-based medical device software in medical physics,” *Phys. Medica*, vol. 83, pp. 1–8, Mar. 2021.
- [52] “General Data Protection Regulation (GDPR) – Official Legal Text.” [Online]. Available: <https://gdpr-info.eu/>. [Accessed: 13-Jun-2021].
- [53] “DocsRoom - European Commission.” [Online]. Available: <https://ec.europa.eu/docsroom/documents/37581>. [Accessed: 13-Jun-2021].
- [54] “ISO - IEC 82304-1:2016 - Health software — Part 1: General requirements for product safety.” [Online]. Available: <https://www.iso.org/standard/59543.html>. [Accessed: 13-Jun-2021].
- [55] “ISO - IEC 62304:2006 - Medical device software — Software life cycle processes.” [Online]. Available: <https://www.iso.org/standard/38421.html>. [Accessed: 13-Jun-2021].
- [56] “ISO - ISO 14971:2019 - Medical devices — Application of risk management to medical devices.” [Online]. Available: <https://www.iso.org/standard/72704.html>. [Accessed: 13-Jun-2021].
- [57] “jreventos/tfm.” [Online]. Available: <https://github.com/jreventos/tfm>. [Accessed: 13-Jun-2021].
- [58] “NVIDIA GeForce RTX 2070 With 8GB vRAM Launching On October 17 For \$499 MSRP (\$599 FE).” [Online]. Available: <https://wccfttech.com/nvidia-geforce-rtx-2070-with-8gb-vram-launching-on-october-17-for-499-msrp/>. [Accessed: 13-Jun-2021].

ANNEX 1

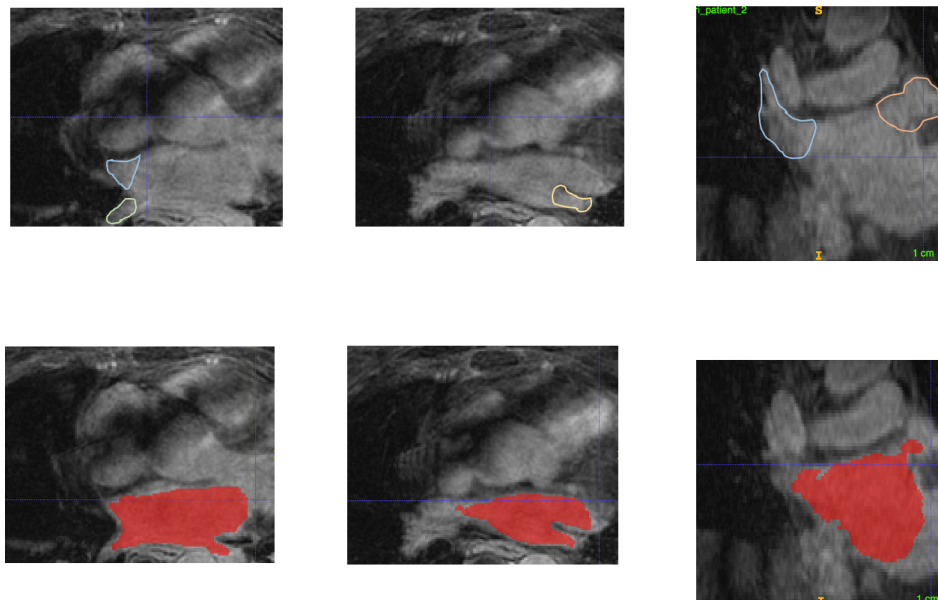
In this annex we attach the PV comparisons between the real PVs and the segmented ones. The colored delineations are only present in the first images, as I got familiar with the anatomy later on and did not have the need to delineate them. The color code for the PVs is the following:

RIPV – Yellow, RSPV – Orange, RCT – Light orange, RMPV – Maroon, LSPV – Green, LIPV – Blue, LCT – Dark blue

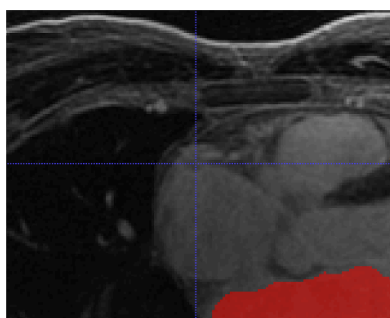
PATIENT 1



PATIENT 2



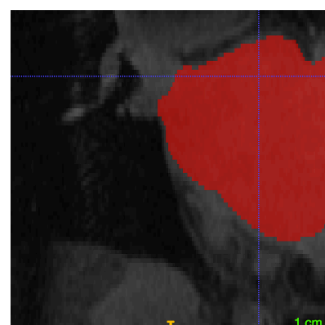
PATIENT 3



Axial

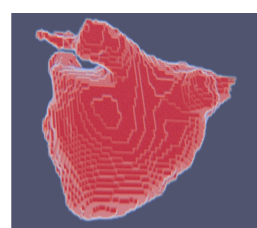
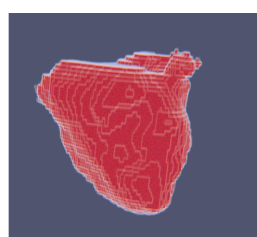
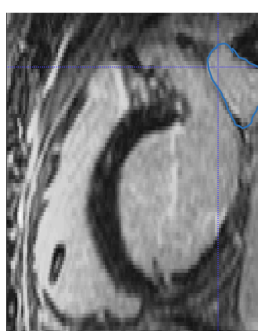
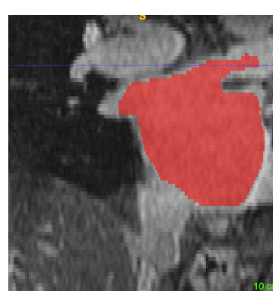
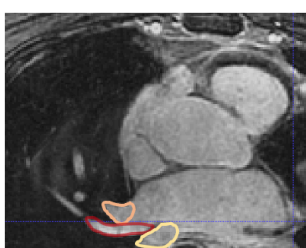


Sagittal

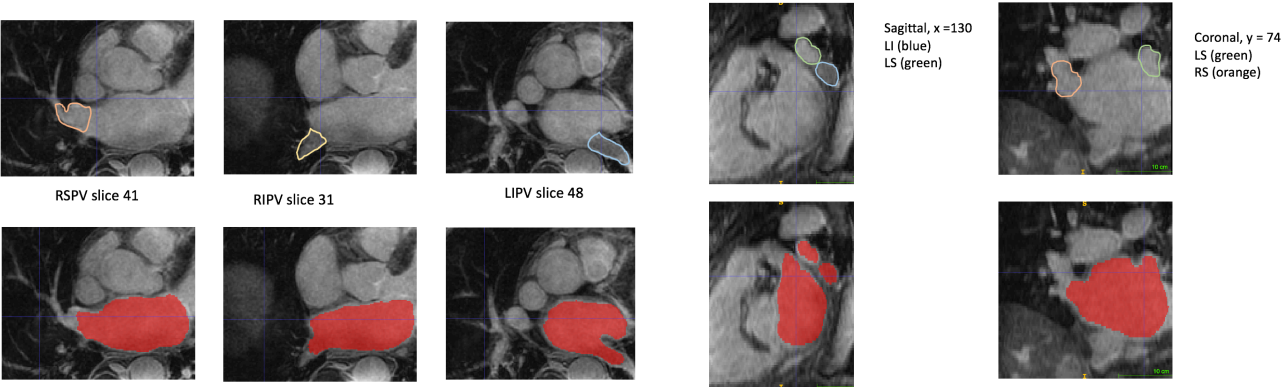


Coronal

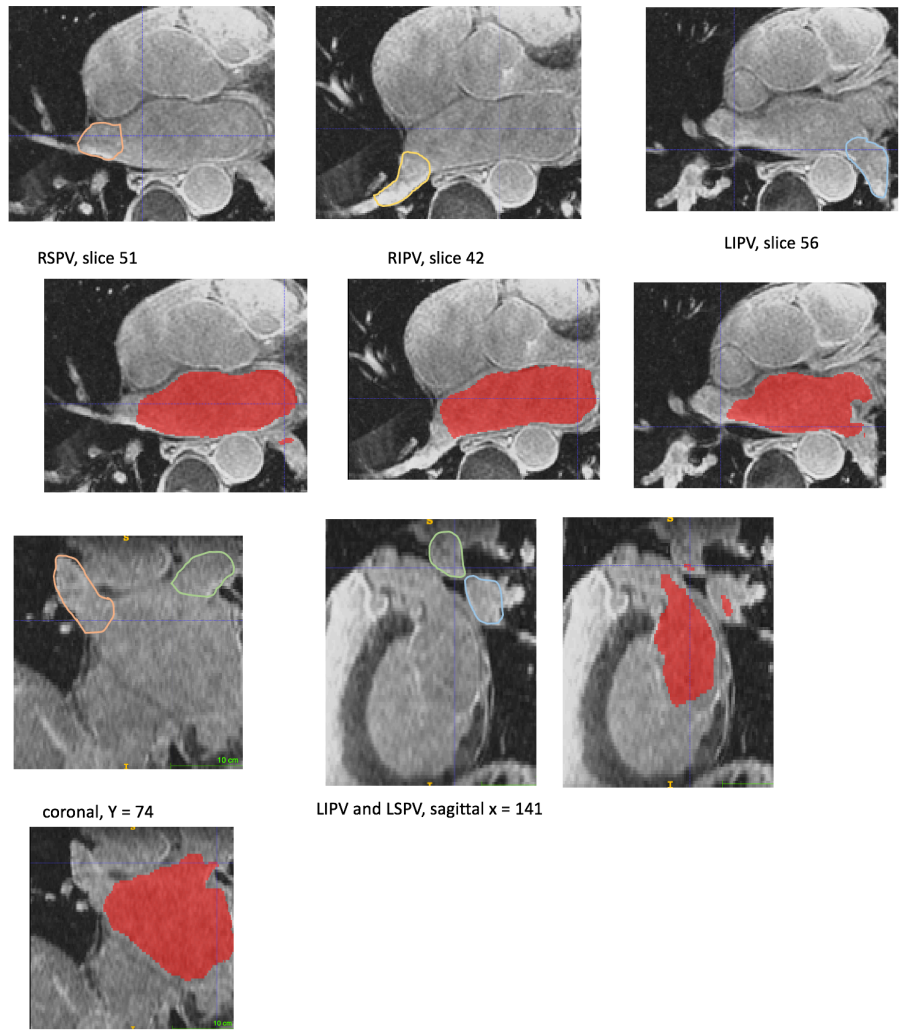
PATIENT 4



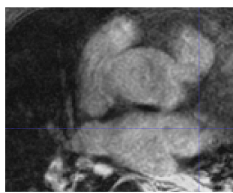
PATIENT 5



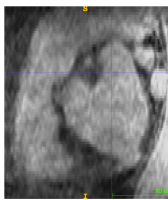
PATIENT 6



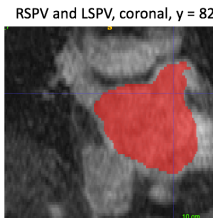
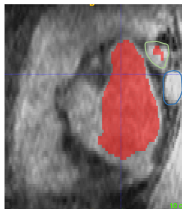
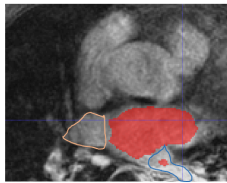
PATIENT 7



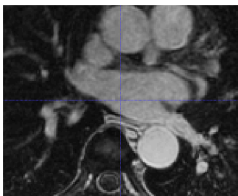
RSPV and LIPV, slice 48



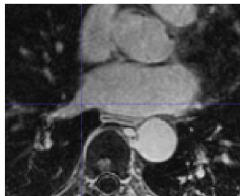
LIPV and LSPV (sagittal, x = 123)



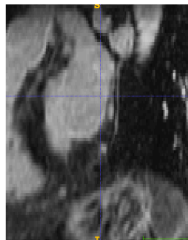
PATIENT 8



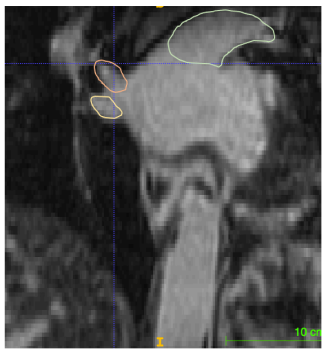
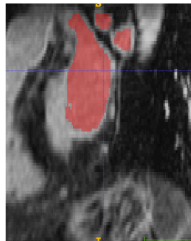
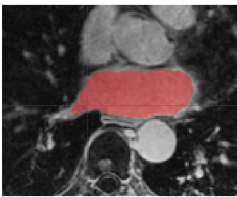
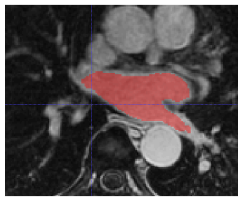
RSPV and LIPV, axial 65



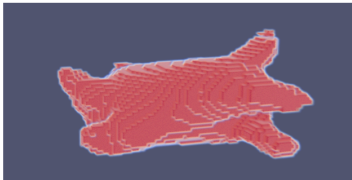
RIPV, axial 57



LSPV and LIPV, sag x = 115

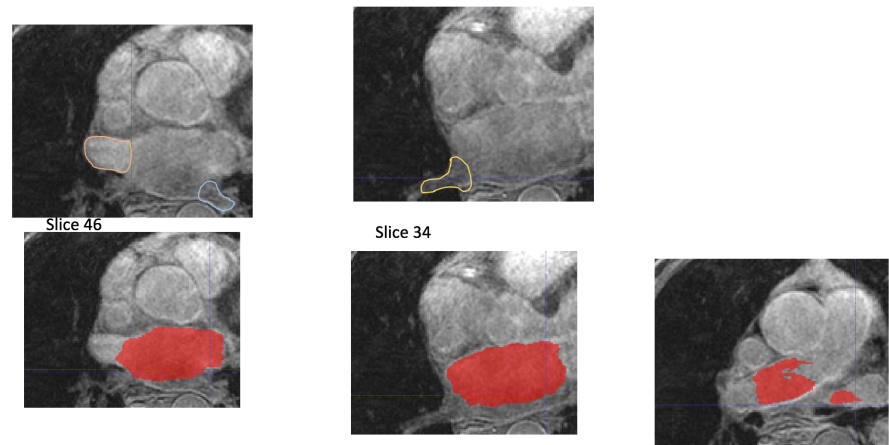


LSPV (green), RSPV (orange), RIPV (yellow)
Coronal, y = 66

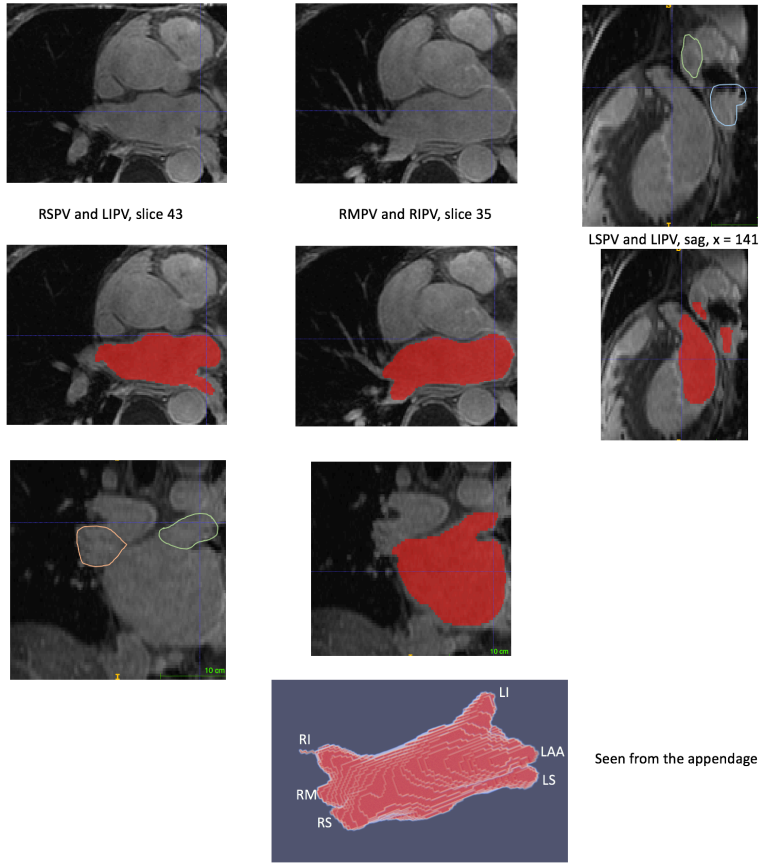


From the top

PATIENT 9



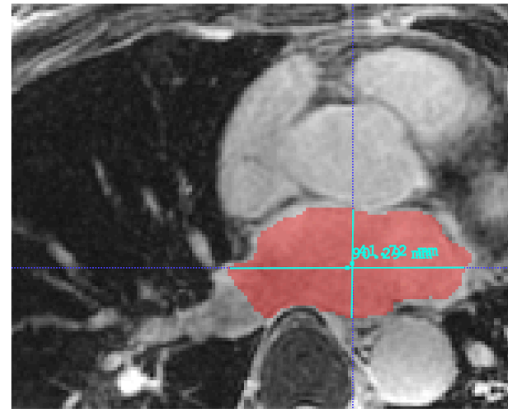
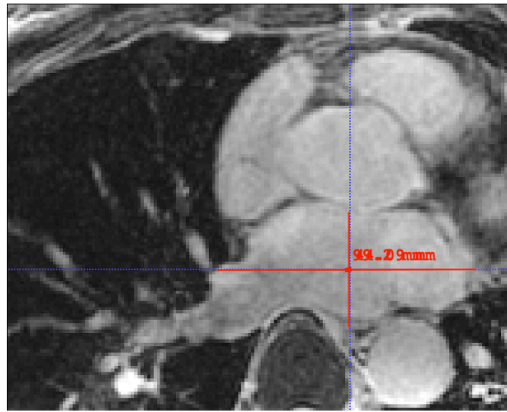
PATIENT 10



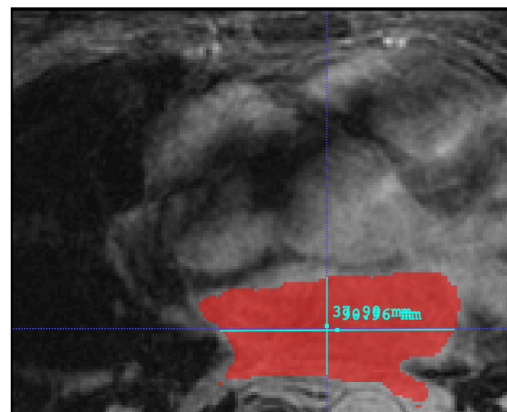
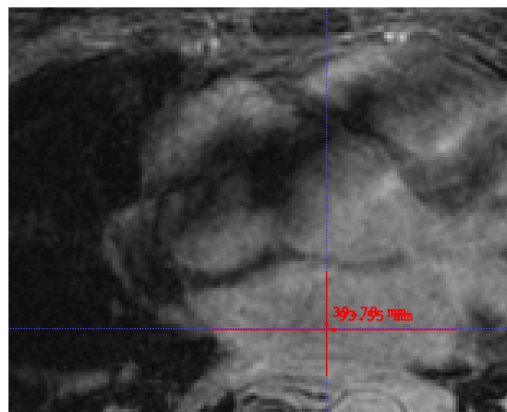
ANNEX 2

In this annex we attach some of the measurements that were made to compare the real LA surface on the axial plane with the segmented one. Not all measures were screenshotted, and that is why some patients are missing:

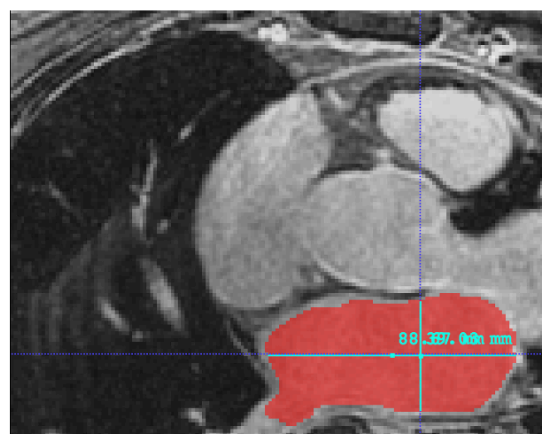
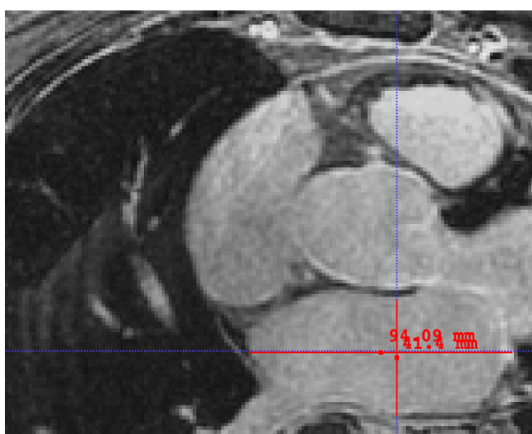
PATIENT 1



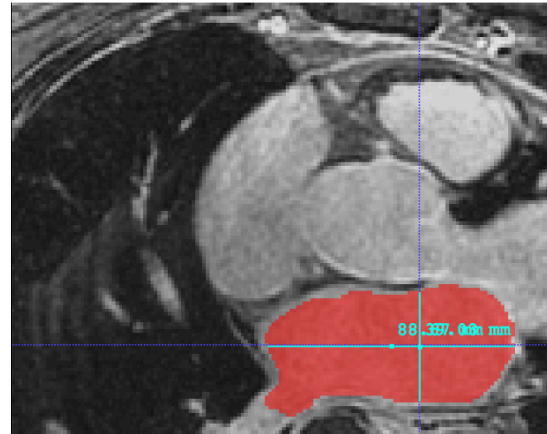
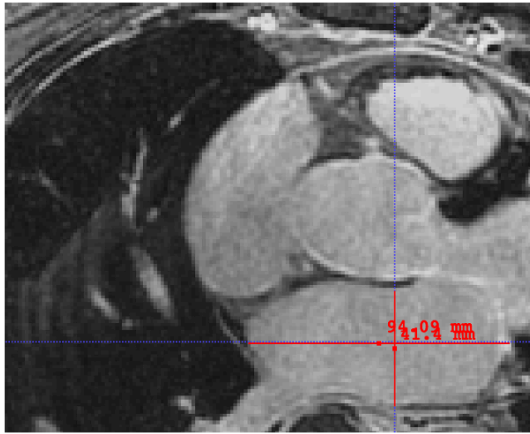
PATIENT 2



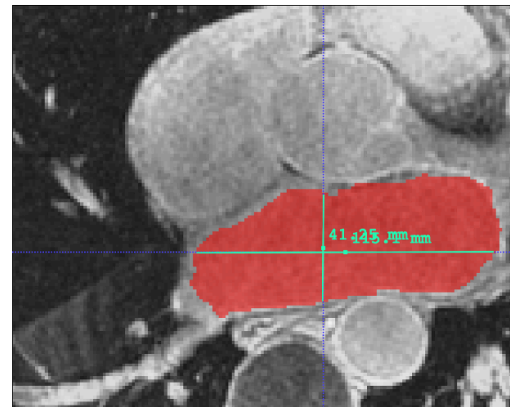
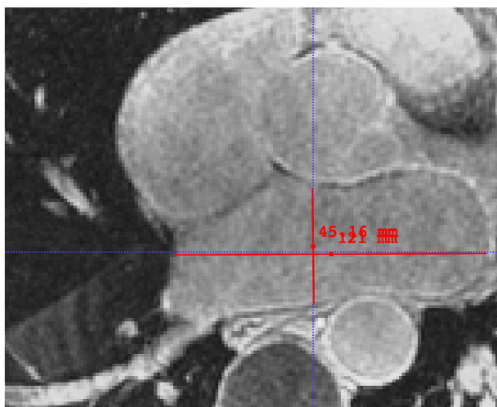
PATIENT 4



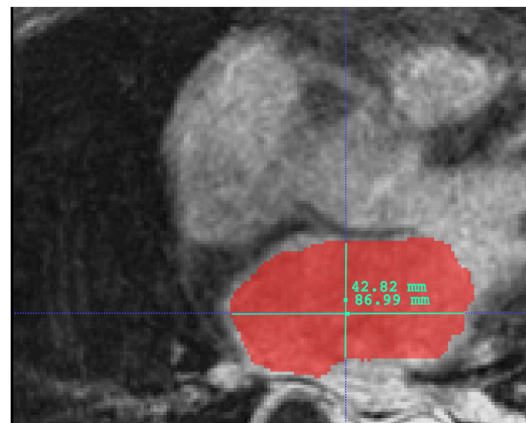
PATIENT 5



PATIENT 6



PATIENT 7



PATIENT 8

